The Production and Perception of Pitch and Glottalization in Yucatec Maya

Melissa Frazier

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Linguistics.

Chapel Hill
2009

Approved by

Jennifer L. Smith
Elliott Moreton
David Mora-Marín
Randall Hendrick
Christopher Wiesen

ABSTRACT

Melissa Frazier: The Production and Perception of Pitch and Glottalization in Yucatec Maya
(Under the direction of Jennifer L. Smith)

This dissertation uses the Bidirectional Stochastic OT model of the phonetics-phonology interface (Boersma 2007a) to analyze the production and perception of pitch and glottalization in Yucatec Maya. The Gradual Learning Algorithm (GLA, Boersma and Hayes 2001) is used to develop mean ranking values of constraints. I show that, when using this algorithm, a simulated learner must be trained on both production and perception tableaus in order to reach an accurate adult grammar (*contra* Boersma 2006, who proposes that perception learning alone is sufficient). This simulated learner is trained on phonetic data obtained from tokens of real speech, and these results show that bidirectional constraints can account for the symmetrical relationship between production and perception. However, because the symmetries are not exact, the production grammar does not simply fall out of perception learning.

Production and perception studies were conducted with native speakers of Yucatec Maya in Yucatan, Mexico. The results of these studies are analyzed with Bidirectional Stochastic OT, but they are also presented in detail in order to document the phonetics of pitch, length, and glottalization in Yucatec Maya. One important result of the production studies is that there is previously undocumented dialectal variation in the production of pitch and length such that tone may be a dialectal feature of Yucatec Maya. Furthermore, there is variation in the perception of pitch that mirrors the variation in production; the cues that differentiate phonemic categories in production are the same cues that are attended to in

perception. These results thus provide further support for the idea that production and perception grammars are defined by the same constraints.

This research fills in two gaps in the literature. First, despite the robust literature on its morphosyntax, there is little research on the sound system of Yucatec Maya, especially at the phonetic level. The production study thus provides the first thorough account of the suprasegmentals of the vowel system, and the perception study is one of the first conducted with this language. Second, this work is the first to test the Bidirectional Model with actual (and not simulated and idealized) language data.

to Chris,
with our heads full of brains
and our shoes full of feet

ACKNOWLEDGEMENTS

Dyos bo'otike'ex uti'al tuláakal:
íicham, ka'ansajo'ob, láak'tsilo'ob, wáa áamigo'ob.

Now that it's all over but the filing, I am very happy to be able to take some time to thank the individuals who have made this dissertation possible.

As my advisor, Jennifer Smith has been as energetic, helpful, and challenging as any grad student could wish for. Jen has a special knack for simultaneously complementing your work and pointing out several ways to improve it; this always made me feel proud of my work and excited about what I had left to do. Thanks, Jen, for all of the insight you have provided through the years, and for all that I'm sure you will provide in the years to come.

I owe Elliott Moreton a particular debt of gratitude for his role in introducing me to acoustic phonetics. When I started grad school, I had never considered focusing on *sound*, but his enthusiasm was contagious and his never-ending encouragement kept me going; I've never looked back. Elliott has also consistently pushed me to think about things from different angles, which, after the initial frustration, tends to lead in productive directions.

Similarly, it was David Mora-Marín who introduced me to Mayan linguistics, and for that I am just as grateful. David has been my mentor in the process of learning to speak Yucatec Maya and in the process of re-learning Spanish. I doubt that either is an enviable position, and so I thank David for his patience and for sharing his vast knowledge of all things Mayan.

Christopher Wiesen has taught me just about everything I know about the applications of statistical analysis. Without his assistance I would never have even considered undertaking this project, and I actually still thank him for that.

As a non-phonologist and non-Mayanist, Randy Hendrick was still able to provide valuable and perceptive comments on various drafts of this work, and I really appreciate the outside perspective he was able to bring. Additionally, I thank Mike Terry for the helpful suggestions he made on my prospectus.

In retrospect, I don't think I could have found a better fit for my grad career than UNC-CH, and I would like to thank the entire department. This includes Professor Emeritus Craig Melchert, who was instrumental in getting me to through my Master's thesis and ready for the Ph.D. level. To my fellow grad cohort – Ian Clayton, Jennifer Renn, and Abby Spears Wright: thanks for joining me in both work and procrastination for the last five years!

I am also grateful for my friends and family. I dedicate this dissertation to my husband, Chris Peters, who not only agreed to marry a grad student but even followed my work closely enough to understand my dissertation defense. Now that's love. My parents, Greg and Donna Frazier, have been nothing but supportive throughout this process, and I truly thank them for that. Finally, my best friends, Chris Leonard and Allison Mastalerz, have done a great job of keeping me in check when I lapse into esoteric linguistic jargon. Girls, I know how long you've been waiting for the day you can officially call me Dr. Magoo. Thanks to all of you for your love, support, and comic relief!

This project could not have been completed without the help of many in the Yucatan. Thanks are especially due to Santiago Domínguez: for insightful language instruction, word list vetting, participant finding, and all-around good conversation. Sharon Mújica may not be a Yucatan resident, but I'll always think of her as one; thanks, Sharon, for helping with contacts and logistics even when I wasn't officially a member of your student group. Thanks also to Fidencio Briceño Chel, Miguel Güémez Pineda, Crisanto Kumal Chan, Ismael May May, and John Tuxill for language instruction, and to all of the participants in the production and perception studies.

This dissertation has been improved by the comments I received from audiences and reviewers at the SSILA Annual Meeting (Chicago, 2008), NELS 39 (Cornell University, 2008), the LSA Annual Meeting (San Francisco, 2009), and WSCLA 14 (Purdue University, 2009). Parts of this dissertation are also discussed in Frazier (2009, in press).

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

LIST OF ABBREVIATIONS

AILLA	Archive of the Indigenous Languages of Latin America (http://ailla.utexas.org)
[cr]	creaky voice
/gl/	GLOTTALIZED vowel
[gs]	glottal stop
/hi/	HIGH TONE vowel
[in]	initial pitch
IPA	International Phonetic Alphabet
[mod]	modal voice
OT	Optimality Theory
ROA	Rutgers Optimality Archive (http://roa.rutgers.edu)
[sp]	pitch span (maximum pitch value minus minimum pitch value)
StOT	Stochastic Optimality Theory
[wg]	weak glottalization

1

INTRODUCTION

This dissertation furthers linguistic research in modeling a phonetic grammar, which includes any language-specific considerations in the production of a continuous output and in the interpretation of that continuous output by the listener. The data under analysis comes from Yucatec Maya, a Mayan language of Mexico (see §1.2), and the model of analysis is Boersma's (2007a) Bidirectional Stochastic OT (see §§1.1.3 and 4.3). The literature on the phonetics-phonology interface has noted that production and perception show many symmetries in the use of acoustic cues (see below and §1.1.1); the cues that differentiate one phonemic category from another in production tend to be the same cues that affect the listener's performance in a discrimination task. These symmetries are found in Yucatec Maya phonetic data (see Chapters 2 and 3), and my analysis (see §5.6) shows that Bidirectional Stochastic OT is able to account for them. However, because the symmetries are not exact, a simulated learner must be trained on both production and perception tableaus in order to reach an accurate adult grammar, which goes against Boersma's (2006) claim that perception learning alone is used to rank the *cue constraints* of his model (see §5.3). Thus, this dissertation shows that the use of bidirectional constraints is appropriate for analysis at the phonetics-phonology interface, but the only way for a simulated language learner to reach an accurate ranking of these bidirectional constraint is for the learner to test both an interim production grammar for its success in accounting for those cues controlled by the speaker and an interim perception grammar for its success in accounting for those cues attended to by the listener.

An example of the symmetrical use of acoustic cues in production and perception comes from the English tense/lax contrast. Escudero and Boersma (2004) summarize a body of literature on the production and perception of this contrast. It is well known that a tense vowel like /i/ is produced with a lower F_1 (i.e. it is higher) and a longer duration than a lax vowel like /ɪ/ (Peterson & Barney 1952, Peterson & Lehiste 1960). The reliability of such cues in distinguishing this contrast (in production) varies by dialect. Escudero and Boersma investigate the effect of dialect variation on the perception of these cues by L1 and L2 users of Scottish and Southern British English. In production, Scottish English speakers distinguish /i/ from /ɪ/ primarily by F_1 and only minimally by duration, whereas duration is the predominant cue to this contrast with Southern British English speakers. The results of Escudero and Boersma indicate that both native and non-native (L1 = Spanish) speakers of the Scottish dialect rely more heavily on spectral cues in perception than native and non-native speakers of the Southern British dialect. Hence the behavior of listeners mirrors the behavior of speakers with regard to the cues that signal the tense/lax contrast.

In Frazier (2006), I document a subtle length difference between two forms that are otherwise homophonous: in American English, dimorphemic words like *passed* are produced with slightly longer vowel durations than monomorphemic words like *past*. Additionally, when performing a forced choice perception task, listeners were more likely to select a dimorphemic word (e.g. *passed*) (as opposed to a monomorphemic "homophone", e.g. *past*) as the vowel length of the stimulus increased. The focus of that paper was a phonological analysis, and so I used phonological forms where vowel duration was a function of the quantity of moras that dominated the vocalic segment (and where mora-sharing accounted for finer categories than just “long” versus “short”). If we shift the perspective to the phonetics-phonology interface, the grammar must account for the symmetrical relationship between the phonological categories (as defined by mora count or some other criteria) and the phonetic dimension of vowel length.

The observation of such symmetries has led to the proposal that production and perception should be handled with the same mechanisms, or specifically in OT phonology, with the same constraints. This is the insight behind Boersma’s (2007a) Bidirectional Model of phonology and phonetics (see §4.3 for further references and discussion). At the phonetics-phonology interface, *cue constraints* relate an acoustic cue to a discrete phonological category in both production and perception. The constraints have the same ranking values in both production and perception and hence account for the noted symmetries.

In this dissertation, I document a variety of details about the production and perception of the suprasegmental properties of vowels in Yucatec Maya. The results of these production and perception experiments indicate that the cues that differentiate phonemic categories in production are the same cues that are attended to in perception. I thus use the Bidirectional Model to account for these symmetries. I will show that this model is capable of accounting for the production and perception of pitch and glottalization in Yucatec Maya. Specifically, cue constraints and articulatory constraints (defined following Boersma 2006, 2007a) can be ranked in such a way to accurately predict how speakers map two different vowel phonemes onto phonetic outputs that are defined for pitch and glottalization and how listeners map these same phonetic forms onto a phonological form. However, when using the Gradual Learning Algorithm (Boersma and Hayes 2001, see §4.3.4) to develop Stochastic OT mean ranking values of constraints, the simulated learner must be trained on both production and perception tableaus in order to reach the most accurate adult grammar (*contra* Boersma 2006, where perception learning alone is used to rank cue constraints, and Boersma and Hamann 2008, where articulatory constraints have fixed universal rankings).

This research fills in two gaps in the literature. First, despite the robust literature on the morphosyntax of Yucatec Maya, there is little research on the sound system of Yucatec Maya, especially at the phonetic level. The production study thus provides the first thorough account of the suprasegmentals of the vowel system, and the perception study is one of the first conducted with this language.[1] For this reason, the results of these studies are presented in detail in order to document the acoustic features of pitch, glottalization, and vowel length in Yucatec Maya. Second, this work is the first to test the Bidirectional Model with actual (and not idealized) language data. In Boersma's own work (see §4.3), he has used data that is simulated to reflect known generalizations about a language but that is not data from actual productions of that language. It is thus important to see how the model works with real language data and all of its irregularities, and I have found that the model can account for the production and perception of pitch and glottalization in Yucatec Maya.

In the rest of this chapter, I present the background discussion and literature review that is necessary to introduce the reader to the phonetics-phonology interface (§1.1) and Yucatec Maya (§1.2). I outline the rest of the dissertation in §1.3.

[1] Avelino (2007) presents the only other perception study I know of with Yucatec Maya speakers.

1.1 Theoretical Background

1.1.1 The Phonetics-Phonology Interface

The goal of phonological theory is to account for all aspects of the *sound grammar*. A majority of work in this area focuses on the relationship between the underlying forms of a language (the forms that native speakers store in their heads) and the surface phonological outputs that speakers actually say. For example, the phonologist may notice that the plural morpheme of English is sometimes pronounced [s], as in *cats*, and sometimes pronounced [z], as in *dogs*, and thus may wish to posit one underlying form for the plural morpheme and account for its different surface forms by using some preferred theoretical device. It is widely accepted, though, that the production of phonological units is not the only part of the sound grammar. For one thing, it has been demonstrated that a phoneme labeled /α/ in one language may not be acoustically identical to a phoneme labeled /α/ in another language. The sound grammar must then also account for language-specific variation at the acoustic (or phonetic) level. For another thing, we also know that the acoustic characteristics of /α/ in one language may not be perceived as /α/ in another language, and thus language-specific perception must be part of the sound grammar.

A well-known example of the language-specific nature of both phonetic production and perception is that of voicing onset time (VOT). In both English and Spanish, voiceless stops contrast with voiced stops. The average VOT (for voiced and voiceless stops, respectively) 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 and Abramson 1964, 1970, Cho and

Ladefoged 1999). Other examples that demonstrate language-specific aspects of phonetic production or perception are discussed throughout this dissertation.

The sound grammar must account for both production (how a speaker uses a stored abstract form to generate an output) and perception (how the listener takes a spoken form to identify a stored abstract form). Not only that, but the sound grammar must also account for the continuous phonetic details of speech. This means that the production grammar must be able to relate (at some level) an abstract categorical form to a continuous form, while the perception grammar must be able to relate that continuous form back to an abstract categorical form. In this dissertation I refer to any categorical form as a *phonological form* and to any continuous form as a *phonetic form*. The part of phonology that deals with the relation between a phonological form and a phonetic form is generally referred to as the phonetics-phonology interface.

Phonological theory has come a long way in its relatively short history of attempting to model the language-specific properties of phonetic (or continuous) forms. The history is short because, for years, the majority opinion of linguistic research was that "phonetics" was not controlled by the mental grammar but was subject to universal constraints on anatomy and physics. The general approach, as exemplified by *SPE* (Chomsky and Halle 1968), was that the phonological component of the grammar produced an output that was categorical but also contained all language-specific information. The actual phonetic output produced by a speaker was thus a universal function that acted on the language-specific phonological form.

Under this theory, one goal of phoneticians was to describe these universal phonetic rules. However, as Keating (1988: 287) points out, "…the more phoneticians looked, the more exceptions they found to possible phonetic generalizations." For example, Ladefoged

(1972) shows that Amharic, Navajo, and Hausa ejectives differ by timing, but timing is not a distinctive feature for these consonants in any one of the languages. In other words, the languages use the same phonological categories, but these categories differ by non-distinctive features. Based on this type of research, it has thus become accepted that *phonetics* is indeed a function of the language-specific grammar, and the phonetics-phonology interface has become and still is a wide area of study. Keating summarizes:

> "Patterns of phonetic detail are interesting, then, not because they constitute a special universal component outside of grammars … but rather because they are an integral part of phonology. It seems likely that there are no true phonetic universals and that the grammar of a language controls all aspects of phonetic form." (Keating 1985b: 129)

Once the idea of a phonetic grammar had gained more support, it was generally assumed that the phonetic grammar was a separate module that was implemented after the phonological module. In Pierrehumbert's (1980) groundbreaking work with English intonation, she uses a phonological grammar that defines the allowable melodies and how they are aligned with metrical structure and a phonetic grammar that associates F_0 values with tones and determines how to fill in the gaps. To defend the need for a phonetic component of the grammar, Pierrehumbert (1980: 11) says, "There is no well-defined level of description which is more concrete than a derived phonological representation, yet still linguistic rather than quantitative in character, at which the linguist may leave off and turn his work over to the physiologist."

In more recent work (e.g. Flemming 2001, Boersma 2007a), the two module approach has been challenged. As will be discussed in more detail in Chapter 4, Flemming proposes that, in production, an underlying phonological form is directly mapped onto a continuous phonetic output, while Boersma proposes that, in production, an underlying phonological form is simultaneously mapped onto a phonological surface form and two distinct continuous forms (one articulatory and one auditory). Thus, in Flemming's model, there is no distinction between phonetics and phonology, and in Boersma's model there are distinct phonological and phonetic outputs but there are not distinct phonological and phonetic modules. Today, there is little agreement among linguists as to how to handle the phonetics-phonology interface (though there is at least general agreement that gradient properties of speech are language-specific). Kingston (2007: 401) sums up the situation: "…the field has reached no consensus about what the interface is, nor has it even agreed that one exists at all."[2]

The model of analysis in this dissertation, Bidirectional StOT states that there is a phonetics-phonology interface and that ranked constraints account for the language-specific relation between a phonological surface form (i.e. not an underlying form) and a phonetic form (what the speaker actually says and what the listener actually hears) in both production and perception. In using Bidirectional StOT to analyze the data from Yucatec Maya, I conclude that this model is advantageous for its use of bidirectional constraints, which account for the symmetrical use of cues by speakers and listeners, and for the stochastic evaluation of constraint rankings, which elegantly accounts for variation. Furthermore, I identify situations in which the use of cues is not exactly symmetrical between speakers and listeners, and hence show that learning through perception tableaus alone (as proposed by Boersma (2006)) will not lead to an accurate adult grammar; the learner must be trained on both production and perception tableaus in order to reach an accurate adult grammar.

[2] To be clear, the researchers that Kingston refers to who claim that no interface "exists at all" are not claiming that phonetics is universal but that there is no distinction between phonetics and phonology.

1.1.2 Optimality Theory

The model of analysis used in this paper is a variation on the framework of Optimality Theory (OT; Prince and Smolensky 1993/2004, McCarthy and Prince 1993, 1995, see McCarthy 2002 for an overview). Though this newer model differs from classic OT in a few crucial ways, it will be useful to briefly summarize some of the main tenets of classic OT here, including a description of how OT analyses are presented with tableaus. The reader who is already familiar with the workings of OT may skip this section.

An OT grammar consists of a language-specific ranking of universal constraints that evaluate the well-formedness of possible outputs. The list of constraints is universal, and OT constraints are violable. A violation of a constraint is tolerated by a language if this leads to the satisfaction of a higher-ranking constraint. In this way, the ranking of constraints creates the grammar of a language. For example, *CODA (syllables do not have codas) is a universal constraint, but not every language bans codas in their outputs.[3] For languages that do not ban codas, some higher ranking constraint must prefer the presence of the coda. In English, an input of /hid/ 'heed' will be mapped onto [hid], which has a coda. This output is optimal because a constraint that bans the deletion of material from the input (MAX) outranks *CODA. In some other language, *CODA will dominate MAX (abbreviated *CODA » MAX) and in this language [hi] would be the optimal output for /hid/.

The optimal output of a language is decided through competition among possible outputs, which are called *candidates*. The constraints assign violation-marks ('*') to any output that violates that constraint. All losing candidates must incur more violation marks from a higher-ranking constraint than the winning candidate (the actual output). In OT, a tableau is used to show the input, the output candidates, the ranking of constraints, and how

[3] In constraint names, the asterisk means 'do not produce such-and-such' or 'such-and-such is banned'.

the violation marks determine the optimal candidate. Some examples are shown in (1.1), in which mini-grammars are defined by the constraints *CODA and MAX (do not delete anything from the input). In (1.1a), we see a language like English. Here the input is /hid/ and the relevant constraints are ranked so that MAX » *CODA. The candidate [hid] incurs a violation of *CODA, but the higher-ranking constraint MAX assigns a fatal violation mark to [hi] (as denoted by the exclamation mark). Thus, [hid] is deemed optimal by this grammar, which is denoted by the pointing finger: ☞. A different type of language is illustrated in (1.1c). In this language *CODA dominates MAX and hence assigns a fatal violation mark to the candidate [hid].

(1.1) example tableaus

a. language that allows codas, input with a coda

/hid/	MAX	*CODA
☞ [hid]		*
[hi]	*!	

b. language that allows codas, input without a coda

/hi/	MAX	*CODA
☞ [hi]		
[hid]		*!

c. language that deletes codas

/hid/	*CODA	MAX
☞ [hi]		*
[hid]	*!	

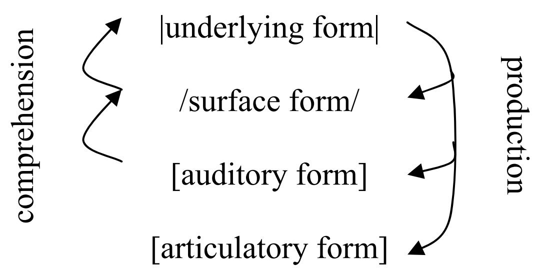
When the violation-marks incurred by the winning output are a subset of the violation-marks incurred by the losing candidate, that losing candidate is *harmonically bounded*. This is the case in (1.1b), where the input has no coda. The candidate with a coda, [hid], is harmonically bounded and thus will never win no matter what the constraint ranking is.

The above discussion has indicated how the ranking order of universal constraints defines an individual language's grammar and how tableaus are used to show the evaluation of a candidate set with respect to some input and some constraint ranking.

1.1.3 Bidirectional Stochastic OT

The specific model of analysis used in this paper is Boersma's (1997, 1998, 2007a,b) Bidirectional Stochastic OT.[4] This model is discussed in detail in §4.3, but a very brief overview is relevant at this time so that we can refer to some of the components of the model before Chapter 4. In the rest of this chapter as well as Chapters 2 and 3, it will be necessary to refer to the types of representations used in the Bidirectional Model and how these representations relate to the tasks of production and perception.

(1.2) Boersma's (2007a: 2031) "bidirectional model of phonology and phonetics"



There are four different representations in (1.2). The topmost two are discrete phonological representations and should be familiar to any phonologist: traditionally, the underlying form is the input to production and the surface form is the phonological output. However, the surface form is not the only output of this model; there are also two continuous phonetic outputs: the auditory and articulatory forms. Note the use of vertical lines (|UF|), slashes (/SF/), and brackets ([AudF] and [ArtF]) to distinguish among the different forms; this notation will be used throughout this paper.

[4] A "Bidirectional Optimality Theory" is also used for analysis in syntax and semantics, where constraints bidirectionally relate form and meaning and hence take into consideration both the speaker and the listener (see Blutner 2000, Zeevat 2000). In this paper I use Bidirectional OT to specifically refer to a theory of phonetics and phonology as developed by Boersma.

In the task of production, the speaker starts with an underlying form as the input. Surface forms, auditory forms, and articulatory forms compete as freely combining candidate triplets. Thus, the optimal candidate set will define a phonological output as well as two phonetic outputs.

The task of comprehension is twofold. In the first stage, called *prelexical perception* (Boersma 2006), the listener starts with the auditory form and, via the grammar, determines the optimal surface form. This surface from is then mapped onto an optimal underlying form to complete the task of comprehension.

The innovation of the model is that the same constraints and the same forms are used in both production and perception. This is a reaction to facts discussed above that show that perception tends to mirror production.

1.2 General Description of Yucatec Maya

In this section I discuss my fieldwork with Yucatec Maya (§1.2.1) and the historical and sociolinguistic context of the language (§1.2.2), and I review the relevant phonological literature (§§1.2.3 – 1.2.8).

1.2.1 Fieldwork

My interactions with communities in Yucatan, Mexico occurred during the summers of 2006, 2007, and 2008. In 2006 I took the first level of the intensive language learning course offered through the University of North Carolina. As part of this course, I spent four weeks in the Yucatan, primarily interacting with residents of Santa Elena and Xocén (/ʃoken/).[5] During this time I did not do any experimentation, but rather observed – and tried

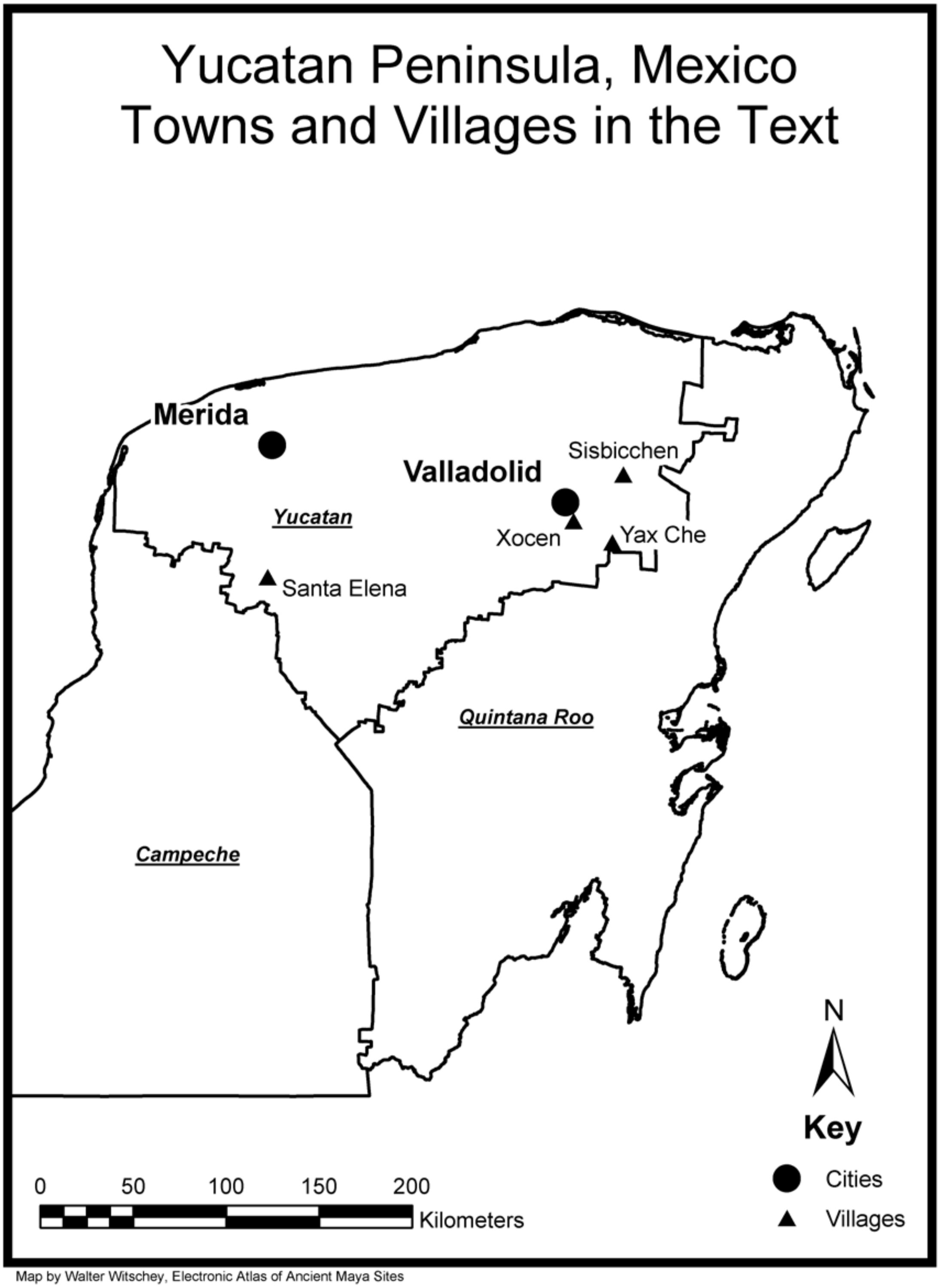
[5] For those interested in the culture of the Yucatan, Xocén is somewhat famous for its contribution to the Caste War and for the religious artifacts that have been and are currently housed there. Terán Contreras and Rasmussen (2005) discuss the vibrant history of this small and traditional village.

to use – the language as much as possible. My notes from this time period are referenced at various places in this dissertation. In 2007, I returned to the Yucatan for six weeks to take the second level of intensive language study. I spent three weeks in Mérida, two in Santa Elena, and one in Sisbicchén (/sisbiktʃen/). Production study 1 was completed at this time. Finally, in 2008, I spent three weeks doing research in Mérida, Santa Elena, Valladolid, and Xocén, and production study 2 was completed at this time.[6] The cities and towns that are referred to in this dissertation are marked on the map in Fig. 1.1.

In 2007 I was also able to make a variety of recordings in which native speakers of Yucatec Maya either told (from memory) or read stories. These recordings are available at the Archive of the Indigenous Languages of Latin America (http://www.ailla.utexas.org) for anyone who would like to listen to the language or use them for phonetic/phonological analysis. Most of the recordings used in production study 2 (see Chapter 2) are also available at AILLA.

[6] Production study 2 includes participants from Sisbicchén, who were recorded in Valladolid, and participants from Yax Che /jaʃ tʃeʔ/, who were recorded in Xocén.

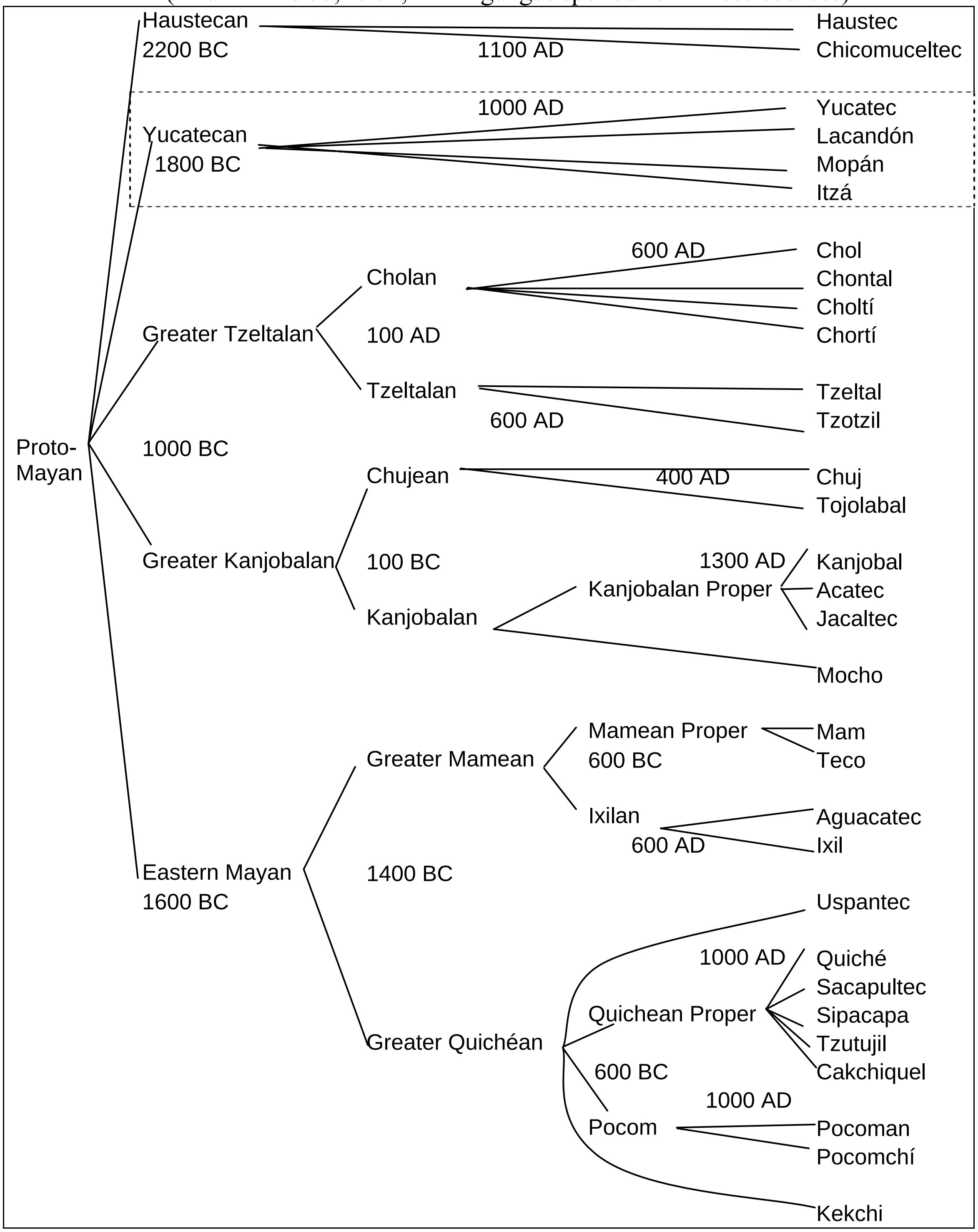
Figure 1.1: Map of the Yucatan Peninsula (Witschey 2009)



1.2.2 Historical and Sociolinguistic Description

Yucatec Maya is spoken by about 700,000 in Yucatan, Campeche, and Quintana Roo, Mexico (according to the *XI Censo de Población y Vivienda* conducted by INEGI, Mexico in 1990; see Güémez Pineda 1994, Briceño Chel 2002, Gordon 2005). Gordon (2005) also claims there are about 5,000 speakers in Belize, but I know of no other source that reports on these speakers. Alternative names and spellings for this language include Yucateco, Yukateko, and Yukatek Maya. Native speakers call their language Maaya' T'aan (literally "Maya words/speech") or simply Maaya'. Academics have consistently used the modifier "Yucatec" to distinguish this language from other Mayan languages, and so I will not depart from this standard. I use the spelling with <c>'s because this is the spelling convention adopted by the communities in Mexico that use and teach Yucatec Maya, as opposed to the spelling with <k>'s, which is a Guatemalan convention (as noted in the "official alphabet for the Guatemalan Mayan languages" (*Acuerdo Gubernativo numero 1046-87* [11/23/87] and its modification *Acuerdo Gubernativo numero 129-88* [3/2/88])). Yucatec Maya is a member of the Yucatecan branch of the Mayan language family (see Fig. 1.2).

Figure 1.2: Mayan languages with estimated dispersion dates
(Kaufman 1976, 1984; all languages spelled as in these sources)



It is debatable where to place Yucatec Maya on the continuum of "language endangerment". In determining this status we wish to consider the number of speakers, the number of children learning the language (especially natively), and the functions of the language in its culture (see Tsunoda 2005 for an overview of perspectives on defining endangerment). My personal experience is limited to the state of Yucatan; there the language is currently thriving in many ways. It is used in daily life by all members of the community except in the larger cities (though I have found that, even in Mérida, you are more likely to be able to speak with a stranger in Yucatec Maya than in English). Its usage is actively encouraged by the government and the non-indigenous citizenry; one of the local TV news programs in Mérida presents a brief summary of the day's headlines in spoken Yucatec Maya. Primary education is mostly bilingual with Yucatec Maya and Spanish. However, secondary education is all in Spanish, and it is rare to find even an older monolingual speaker.

Krauss (1992) distinguishes among *moribund* languages (those not spoken by children), *endangered* languages (those for which transmission to children will cease in the 21st century), and *safe* languages (those which will be spoken by children in the 22nd century). I would argue that Yucatec Maya is safe by this definition, but there are so many variables with regard to the functions of Spanish in the community that it would be foolish for linguists to assume that Yucatec Maya will always be around. This uncertainty about Yucatec Maya's future and the fact that this language is understudied in the areas of phonetics and phonology make phonetic documentation of Yucatec Maya an important area of linguistic research.

1.2.3 Consonants

In this and the next subsections, I introduce the relevant aspects of the phonology of Yucatec Maya. This includes a discussion of the phonemic contrasts, stress patterns, and phonotactics of the language, with special attention paid to the suprasegmental contrasts of length, pitch, and glottalization in the vowel system. Those interested in a more comprehensive grammatical overview should consult Bricker et al. 1998 (henceforth B et al.) and Blair and Vermont Salas 1965 (henceforth B&VS). Unless other citations are given, the details in these sections come from these sources. Some observations (as noted) come from my own field work experience (see previous section).

The consonants of Yucatec Maya are shown in Table 1.1. Yucatec Maya uses many different types of voiceless obstruents – stops, affricates, fricatives and ejectives – but only one voiced obstruent – the implosive [ɓ]. The glottals – [ʔ, h] – are contrastive in both onset and coda position. All sonorants are voiced.

Table 1.1: Consonantal phonemes of Yucatec Maya

		labial	alveolar	post-alveolar	palatal	velar	glottal
stops	voiceless	p	t			k	ʔ
	ejective	p’	t’			k’	
	implosive	ɓ					
affricates	voiceless		ts	tʃ			
	ejective		ts’	tʃ’			
fricatives			s	ʃ			h
nasals		m	n				
lateral approximants			l				
approximants		w			j		

1.2.4 Vowels

The vowel inventory, shown in Table 1.2, is interesting in that there are only the five canonical vowel qualities – [i, e, a, o, u] – but each vowel quality can occur with four contrastive sets of suprasegmental features, resulting in 20 unique possible syllable heads. I refer to each contrastive combination of suprasegmental features (length, pitch, and glottalization) as a *vowel shape*.[7] Throughout this paper I denote each vowel shape with small capital letters: SHORT, LOW TONE, HIGH TONE, GLOTTALIZED. This convention is used so that references to a generic phonological or phonetic property, such as high tone or glottalization, will not be confused with references to a specific phonemic category of Yucatec Maya (i.e. HIGH TONE or GLOTTALIZED) The contrastive nature of vowel shape is demonstrated by the minimal quadruplet below.

Table 1.2: Vocalic phonemes of Yucatec Maya

quality:		shape (applies to any vowel quality):	
i	u	v	SHORT: short, unmarked for tone, modal voice
e	o	v̀v	LOW TONE: long, low tone, modal voice
	a	v́v	HIGH TONE: long, high tone, modal voice
		v́v̰	GLOTTALIZED: long, high tone followed by creaky voice

(1.3) minimal quadruplet for vowel shape:

SHORT	*chak*	‘red’	[tʃak]
LOW TONE	*chaak*	‘boil’	[tʃàak]
HIGH TONE	*cháak*	‘rain’	[tʃáak]
GLOTTALIZED	*cha’ak*	‘starch’	[tʃáa̰k]

The vowel shapes have been described in detail in the literature (e.g. Pike 1946, Blair 1964, Fisher 1973, 1976, McQuown 1975, B&VS, B et al.), and there is general agreement about the descriptions of pitch contours and glottalization. The SHORT vowel is described as

[7] I am not aware of a consistent term in the literature for vowel contrasts that consist of some combination of tone, phonation type, and/or nasalization. It does seem that these suprasegmental properties pattern in similar ways across language families, e.g. Acoma uses a “glottal accent” that is described in a similar manner as Yucatec Maya’s GLOTTALIZED vowel (Miller 1965). It would thus be beneficial to have a shorthand way to refer to these properties (in the same way that *vowel color* is used to refer to combinations of backness and rounding features). The term *vowel complexity* has been used to refer to combinations of tone and phonation type (e.g. Roffe 1946, see also Silverman’s (1997) use of *laryngeal complexity*), but this term has also been used to describe the number of vowel sounds that make up a phonemic unit (i.e. to distinguish monophthongs, diphthongs, and triphthongs). I thus propose *vowel shape* to fill in this terminology gap.

having short duration and neutral tone (meaning that it does not participate in a tonal contrast, B&VS: 7), while the LOW TONE vowel is described as having long duration and low and level pitch (B&VS: 9, Pike 1946: 84).

Accounts of HIGH TONE and GLOTTALIZED vowels are more detailed. B&VS claim that HIGH TONE has different realizations depending on the position of the syllable that bears high tone in the word: HIGH TONE vowels in word-final syllables (including the only syllable of monosyllabic words) start with high pitch which falls throughout the duration of the vowel, whereas vowels in non-final syllables do not fall in pitch in the same way. Gussenhoven and Teeuw (2008) demonstrate that HIGH TONE vowels have different pitch contours depending on the position within the phrase: phrase-final HIGH TONE vowels have a falling contour, while all others have a rising contour. In §§2.4.3 and 2.5.3, I present further evidence in support of the effect of phrase position on pitch contours of HIGH TONE vowels, but as discussed in Frazier (2009), I have not found conclusive evidence in support of the effect of word position on pitch contours.

The GLOTTALIZED vowels are unanimously described as being produced with initial high pitch and modal voice that is interrupted by a glottal stop (i.e. [v́ʔv]), though B&VS and B et al. recognized that this glottal stop does not always have a full realization, especially in rapid speech.

Until recently, descriptions of the vowel shapes were done by ear. Fisher is a notable exception in that he graphically displays the pitch contours of productions by a native speaker, but he does not conduct controlled experimentation or statistical analysis. Several recent sources have begun the important task of acoustic documentation (Avelino et al. 2007, Kügler et al. 2007, Gussenhoven and Teeuw 2008), but each of these papers has focused on either a subset of the vowel shapes or a subset of the contrastive suprasegmental properties. It is of course necessary to test the descriptions of the older literature with phonetic studies. For example, as noted by Avelino et al. 2007 and as will be shown in Chapter 2 (see especially §2.4.2 and 2.5.2), the GLOTTALIZED vowels are rarely produced with a full glottal stop, and instead are generally produced with creaky voice. For this reason, Table 1.2 provides a description and phonological notation – [v́v̰] – that reflects this reality.

In Boersma's model of phonetics and phonology (see §4.3), there are two different levels of phonological representation and two different phonetic forms. Because this is the model of analysis in this paper, it will be helpful to understand how the phonological claims about Yucatec Maya vowel shape might be understood in terms of this multi-level model. The vowel shape representations in Table 1.2 are clearly phonological forms. If we treat these as *surface forms* – the outputs of the phonology, which are then mapped onto phonetic forms – we could easily conceive of a set of markedness constraints that would, if ranked high enough, rule out all other competing surface forms no matter what underlying form is used as the input. I thus also intend for these forms to be the underlying forms we can presume to have been created following lexicon optimization (Prince and Smolensky 1993/2004: 225). We should keep in mind, though, that the surface forms will differ in different syntactic contexts due to the demands of the intonational phrase.

1.2.5 Differences between Orthography and IPA

Before moving on to the other relevant phonological aspects of Yucatec Maya, a brief note on the orthography of the language is necessary. When discussing the actual sounds of the language (whether on a phonological or phonetic level), I will continue to use the IPA,[8] but when presenting example words and phrases, I will use the orthography of the language as employed by those who use and teach Yucatec Maya in Mexico. Differences between IPA and orthography are shown in (1.4).

(1.4) correlations between IPA and Yucatec Maya orthography

consonants:

[ɓ] = <b>; [ʔ] = <’>; [tʃ(’)] = <ch(’)>; [ʃ] = <x>; [h] = <j>; [j] = <y>

vowel shapes:

SHORT = <v>; LOW TONE = <vv>; HIGH TONE = <v́v>; GLOTTALIZED = <v’v>

Note, also, that word-initial glottal stops are not represented orthographically. Any vowel-initial spelling denotes a word that begins with a glottal stop, e.g. *ich* ‘eye’ [ʔitʃ] (see §1.2.7).

1.2.6 Stress

Stress is noncontrastive in Yucatec Maya. However, there are no comprehensive studies of stress, and there are some disagreements in the literature about exactly where stress occurs. B et al. say

> “In words of more than one syllable, the syllable containing a long vowel is stressed. If the word has two long vowels, the vowel in the first syllable receives primary stress. If the word has no long vowel, the last syllable is stressed.” (Bricker et al. 1998: xiii)

B&VS agree that long vowels attract stress but claim that, in disyllabic words with two short syllables, both syllables have equal intensity. Blair (1964) says that, in strings of short, open syllables, the first syllable and then every other syllable after that receive stress, except for sequences of two syllables only, in which case both syllables have equal prominence.

[8] I use IPA instead of the Mayan transcription system because it is important for work on Mayan languages to be accessible to all linguists and not just Mayan specialists. For this reason, the transcriptions in this paper may differ from those in the sources cited if those sources use the Mayan system.

Generalizing from the above statements, we can conclude that heavy syllables and the leftmost syllable attract stress in Yucatec Maya. We could also say that the rightmost syllable attracts stress, but because this syllable is always closed and thus (at least) bimoraic (see next section), this fact is already accounted for by the previous statement.[9] It is unclear from the literature how any stress clashes that might result from these guidelines would be resolved. It would be beneficial for future studies to investigate stress patterns in Yucatec Maya, including the interaction of stress with weight, tone, and intonation.

1.2.7 Phonotactics

Almost all roots in Yucatec Maya are of the form CVC or CVCVC. Prefixes tend to be vowel final and suffixes vowel initial, such that most syllables (in mono- or polymorphemic words) are open and have onsets, except for the final syllable of a word, which is closed. Words are prohibited from beginning or ending with a vowel, as evidenced by the treatment of vowel-initial and vowel-final loan words from Spanish. A glottal stop is epenthesized at the beginning of loan words that begin with a vowel, and an [h] is epenthesized at the end of loan words that end with a vowel:[10] e.g. *amigo* ‘friend’ > [ʔáamigóoh].[11,12]

All consonants can occur in any position in the syllable/word, though there are some co-occurrence restrictions and non-obligatory neutralization. In monosyllabic words of the

---

[9] It is generally assumed that coda consonants are moraic in Yucatec Maya. See Krämer (2001) for an analysis that relies on this assumption.

[10] See AnderBois (in press) for an OT analysis of phrase-final [h]-insertion in Yucatec Maya.

[11] Certain phones exist only in Spanish loans, and hence were not provided as phonemes in Table 1.1. These include [b, d, g, f, ɲ, r], but vary by region. For example, the speakers I have contact with in Xocén produce a [p] in place of [f] in Spanish loans (though these same speakers use an [f] when speaking Spanish), whereas the speakers I know in Santa Elena use [f] when speaking both Yucatec Maya and Spanish.

[12] According to Hanks (1984), unstressed (or secondarily stressed) syllables in Spanish loans receive high tone in Yucatec Maya.

form C_1VC_2, if C_1 and C_2 are both ejectives or if they are both voiceless obstruents with the same place of articulation, they must be identical in all respects (e.g. *k'i'ik'* /k'ḭ́ik'/ 'blood', */kḭ́ik'/, */t'ḭ́ik'/; *peek'* /pèek'/ 'dog', */p'èek'/, */pèep'/). Place of articulation in nasals is noncontrastive in coda position. For example, *k'áan* 'hammock' can be produced as [k'áan], [k'áam], or [k'áaŋ], with [k'áam] being the most frequent production that I hear.

1.2.8 Grammatical Use of Vowel Shape

There are a variety of morphosyntactic processes that can alter underlying vowel shape in Yucatec Maya (see B et al, Blair 1964, Briceño Chel 2006, Ayers and Pfeiler 1997). This is especially prevalent in the verbal system. For a large set of verbs called *root transitives* (as opposed to the *derived transitives*), the active form of the verb has a SHORT vowel (e.g. *kin p'ejik* 'I chip it.'),[13] the agentive passive has a GLOTTALIZED vowel (e.g. *ku p'e'ejel tumeen ten* 'It is chipped by me.'), the antipassive has a LOW TONE vowel (e.g. *kin p'eej* 'I chipped.'), and the agentless passive ("middle voice") has a HIGH TONE vowel (e.g. *ku p'éejel* 'It gets chipped.') (B et al: 333). However, Avelino et al. (2007) present data from nine native speakers of Yucatec Maya which shows that, for the antipassive and middle voice, the F_0 contours of the nucleus of the verb are determined by underlying vowel shape and not by grammatical voice.

Additionally, possession of a noun can trigger a change in vowel shape. For example, *k'áan* 'hammock' is produced with low tone when it is possessed: *in k'aan* [iŋk'àan] 'my hammock' (Lehmann 1998). This process is sporadic in that it does not affect all nouns with

[13] There are also root transitives for which the underlying vowel shape is not a SHORT vowel, and hence the active form does not have a SHORT vowel, e.g. *a'al* 'speak', *na'at* 'understand', síit' 'jump' (Briceño Chel 2006).

an underlying HIGH TONE vowel (e.g. *púuts'* 'needle'; *in púuts'* 'my needle'), and it would be beneficial for future work to investigate this phenomenon in more detail.

Though it is clear that more research is needed on the effect of grammatical properties on the production of vowel shape, the fact that underlying vowel shape can be altered by various grammatical processes is taken into consideration in the design of the production and perception experiments that are presented in Chapters 2 and 3.

1.3 Outline of Dissertation

The chapters of this dissertation are organized as follows. Briefly, Chapters 2 and 3 present the results of production and perception studies with native Yucatec Maya speakers, Chapter 4 discusses some theoretical approaches to the phonetics-phonology interface and explains in detail Bidirectional Stochastic OT, Chapter 5 uses Bidirectional StOT to analyze pitch and glottalization in the HIGH TONE and GLOTTALIZED vowels of Yucatec Maya, Chapter 6 discusses some of the implications for future work with Bidirectional StOT and the GLA that came out of the analyses of the previous chapter, and Chapter 7 presents the conclusions of this dissertation.

In Chapter 2, the methodology and results of two production studies are presented. Participants from Mérida, Santa Elena, Sisbicchén, Xocén, and Yax Che were recorded as they read target words in both isolation and frame sentences. The goal of these production studies is to document the suprasegmental properties of vowel shape in Yucatec Maya. The results indicate that speakers from Mérida and Santa Elena produce vowel shape mostly as it is described in the literature (§1.2.4), but that participants from Sisbicchén, Xocén, and Yax Che do not. I propose that this evidence indicates a dialectal split between the western and eastern parts of the Yucatan Peninsula. This chapter also introduces some new methodology for the measurement of pitch and discusses how pitch and creaky voice interact in different ways for male and female speakers.

The methodology and results of a perception study are presented in Chapter 3. This study was designed to determine how the cues of pitch and glottalization are used to discriminate HIGH TONE and GLOTTALIZED vowels. Participants performed two different tasks, one with natural stimuli and one with manipulated stimuli. The results of the task with natural stimuli indicate that listeners use all of the cues that are differentiated in production in making their decision about which sound they heard. This means that listeners from different dialect groups listen for different cues. The results of the task with manipulated stimuli were different; in this task participants used glottalization alone in making their decision. I discuss how the difference between natural and manipulated stimuli could account for the different results of the two tasks. I propose that the results of the task with manipulated stimuli are still language-specific; i.e., they are dictated by the Yucatec Maya grammar, and that the perception grammar is adjusted in the face of degraded language situations to focus on the cue that is known to be the most successful at distinguishing between two competing phonemes.

In Chapter 4, I present background literature on the phonetics-phonology interface and discuss some of the models that have been proposed in the literature to account for this interface, including Keating's (1990) Window Model and Flemming's (2001) Unified Model. I then introduce the model of analysis used in this paper, Boersma's (2007a) Bidirectional StOT Model. I discuss the properties of this model, how the GLA is used to develop StOT rankings, how PRAAT is used to simulate a GLA learner, and the research that has been done with Bidirectional StOT.

The analysis of pitch and glottalization in Yucatec Maya using Bidirectional StOT is presented in Chapter 5. In this chapter I review the relevant data on production and perception of the HIGH TONE and GLOTTALIZED vowels by speakers from Santa Elena and then introduce the relevant cue and articulatory constraints for the model. These constraints are based primarily on Boersma's proposals about Bidirectional StOT constraints, but previous proposals in the OT literature are also taken into consideration. Some articulatory constraints account for cross-linguistic facts about the interaction of pitch and glottalization which have not yet been analyzed in terms of effort (the usual criteria for an articulatory constraint). Given the particular constraint set and the known distributions of surface form – phonetic form pairings from the production studies (Chapter 2), the GLA is used to learn an adult grammar that can account for the production and perception of pitch and glottalization with the HIGH TONE and GLOTTALIZED vowels. A learning simulation in line with lexicon-driven perceptual learning (Boersma 2006) is first run, and the resulting grammar is highly successful in terms of perception but cannot account for some patterns in the production data. In order to improve this aspect of the grammar, a new learning simulation is run in which the learner is trained on both production and perception tableaus. This new simulation leads to a better production grammar. The conclusion of this chapter is that an accurate Bidirectional StOT grammar can be learned with the GLA when the training data comes from tokens of actual speech but only if the learner is trained on both perception and production tableaus. Both production and perception learning are necessary to learn all the language-specific patterns in the production data.

In Chapter 6, I address some remaining questions about the implementation of the Bidirectional Model and about GLA learning, including the use of finely grained cue constraints and the differences between the evaluation of multiple phonetic dimensions at once versus each dimension separately in production and perception. I then discuss the synchronic and diachronic predictions that are made by this model and my proposal that the learner must be trained on both production and perception tableaus.

Conclusions are presented in Chapter 7. I discuss how the production and perception studies have contributed to our understanding of the phonetic properties of Yucatec Maya. This includes the topics of dialect and idiolect variation and of the nature of Yucatec Maya as laryngeally complex. I then summarize the grammar of Yucatec Maya that we have developed and assess the success of Bidirectional StOT and the GLA for use with real language data.

2

PHONETIC DESCRIPTION OF YUCATEC MAYA: PRODUCTION OF LENGTH, PITCH, AND GLOTTALIZATION

In this chapter, I present the results of two production studies designed to provide details on vowel length, pitch, and glottalization in the vowel system of Yucatec Maya. In §1.2.4, we saw descriptions of the contrasting vowel shapes. As discussed at that time, these descriptions give us a general picture of what suprasegmental features contribute to the contrasts, but they do not provide language-specific acoustic details about the production of these vowels. For example, we do not know how much longer long vowels are than SHORT vowels, if the high tone of GLOTTALIZED vowels is different from the high tone of HIGH TONE vowels, how pitch is produced with the SHORT vowels, nor the characteristics of glottalization with the GLOTTALIZED vowels. In this section we will answer these questions while developing a thorough picture of what values for length, pitch, and glottalization are acceptable for each of the vowel shapes. The ultimate goal is to determine how the vowel shapes are differentiated in production so that we can test for symmetries between production and perception, which will be the topic of the next chapter. The phonetic descriptions of vowel shape that are developed in this chapter will be used to motivate perception experiments in Chapter 3, and the results of this and the next chapter will form the basis of the theoretical analysis of Chapter 5.

In production study 1, participants produced target words in isolation (no frame sentence was used). The goal of this study is to analyze the citation form of the four vowel shapes. In production study 2, participants produced target words in four different frame sentences, so that the production of vowel shape as conditioned by prosodic context could be measured. I will thus identify any effects of prosodic context on vowel shape, but it is beyond the scope of this dissertation to analyze why prosodic context has the effects it does.

In addition to providing the first thorough acoustic analysis of vowel shape in Yucatec Maya, the production studies also point to some dialectal and idiolectal variation, and these results will also be discussed in this chapter. Because of the dialectal variation, the theoretical component of this dissertation (Chapter 5) will focus on data from speakers from a single place. I use speakers from Santa Elena as my focus for two reasons: 1) they produce the acoustic features of vowel shape in a manner that is mostly consistent with phonological descriptions from previous literature and 2) I was able to record enough speakers from this town (12 in production study 1 and 14 in production study 2) to perform meaningful statistical analysis with just their results.

This chapter is organized as follows. I first discuss the methodologies used for each study in §2.1. The participants and procedures of production study 1 are discussed in §2.1.1, followed by the details of production study 2 in §2.1.2. In §2.1.3 I describe how measurements were extracted from the recordings, which are relevant to both production studies. In evaluating the procedures used for the measurement of pitch, we will discover an interaction between pitch and creaky voice that cannot be ignored. This interaction is discussed in detail in §2.2, and the methodologies of the production studies are summarized in §2.3. The rest of the chapter is devoted to the results of the production studies. I present the results of both studies for speakers from Santa Elena in §2.4. In order to explore the dialect variation documented by these studies, results with regard to all other speakers are presented in §2.5. The data presented in this section will not be used for theoretical analysis, and it is provided in order to document this newly discovered dialect variation. I provide local summaries in §2.4.4 (for Santa Elena speakers) and §2.5.4 (for all other speakers) and a chapter summary in §2.6. The goal of §§2.4 and 2.5 is to document the phonetic details of vowel shape in Yucatec Maya and to show exactly which suprasegmental properties – length, pitch, glottalization – contribute to contrast when comparing any two vowel shapes. These sections are rather heavy on the details, and the reader more interested in theory and less interested in phonetic documentation may read only the local and chapter summaries and proceed without loss of relevant information.

2.1 Methodology

This section includes information about the participants and the full details of the procedures used for each production study separately. I then discuss the methods used to extract data on vowel length, pitch, and glottalization, which are the same for both production studies.

2.1.1 Participants and Procedures for Production Study 1 (Isolation)

In total, 24 native speakers participated in this study (see Table 2.1), which took place in June and July 2007. Participants lived in Santa Elena, Mérida, and Sisbicchén, Yucatan, Mexico (see map in Fig. 1.1). The methodology and results of this study are also extensively discussed in Frazier (2009).

Table 2.1: Participant data (production study 1)

		ages	additional languages
Mérida	6 males	33, 39, 40, 41, 47, 47	fluent in Spanish, one also proficient in English
	1 female	39	
Santa Elena	5 males	22, 25, 43, 63, 68	fluent in Spanish, two also fluent in English
	7 females	19, 20, 25, 30, 33, 35, 63	
Sisbicchén	2 males	30, 41	fluent in Spanish
	3 females	24, 29, 30	understand Spanish, but do not use it

Participants were recorded while they read 100 words of the form C_1VC_2 (where C_1 and C_2 were sometimes identical). There were a few exceptions to the CVC pattern. Because speakers in Santa Elena (and surrounding areas) do not produce the implosive [ɓ] in coda position, the word list for speakers from Santa Elena was amended such that the 20 words ending in the implosive were affixed with a vowel initial suffix (if possible; otherwise a nonce form was used).[14] Thus, the form CVɓ becomes CVɓVC, where measurements are taken from the first vowel in the context /CVɓ/. Additionally, both the word list for Santa Elena and the general word list used the word *k'aaba'* 'name', where measurements were taken from the final syllable /ɓaʔ/.[15]

The word list was balanced on the basis of vowel shape and consonant type. The consonant types were glottal stop [ʔ], ejective [p', t', k', ts', tʃ'], implosive [ɓ], voiceless

---

[14] Another goal of production study 1 was to measure the effect of consonant type on the production of pitch. The results of this analysis are presented in Frazier (2009) but are not discussed here. In order to facilitate this goal, the word list was balanced on the basis of consonant type. For this reason, it was important to make sure participants would pronounce the bilabial implosive and not a glottal stop, and it was also necessary to include nonce forms, as discussed in the next paragraph.

[15] As mentioned in §1.2.4, there are claims in the literature that the production of pitch with the HIGH TONE vowel varies depending on the position of the syllable in the word. However, the results of this production study, as explained in Frazier (2009), did not find any conclusive evidence that pitch contours were affected by syllable position within the word. For this reason, syllable position is ignored when the results of production study 1 are presented in §§2.4 and 2.5.

pulmonic obstruent [p, t, k, ts, tʃ, s, ʃ], and voiced sonorant [m, n, l, w, j]. All four vowel shapes were used. Each word matched one possible combination of C_1 type, V shape, and C_2 type (5 x 4 x 5), resulting in 100 words. For ease of exposition, I will henceforth refer to C_1 as the onset consonant and C_2 as the coda consonant (even though not all C_2's are actually codas, due to some polysyllabic target words).

Though all 100 forms required to complete the word list utilize legal phonotactics, there are lexical gaps that were filled in with nonce forms. For example, there is no existing monosyllabic Yucatec Maya word of the form glottal stop - GLOTTALIZED vowel - glottal stop, so the nonce form *a'a'* /ʔáa̰ʔ/ is used.

When more than one existing word of a certain form could be found, the word used in the study was chosen on the basis of the following criteria (in order of importance): grammatical category (nouns and adjectives preferred, then nonclitic function words (e.g. *ich* 'in', *bíin* '*future aspect*'); verbs avoided due to grammatical use of vowel shape (see §1.2.8))[16], familiarity of word (highly familiar words, as rated by a native speaker, chosen over less familiar words), consonant type (stops chosen over affricates (for both ejectives and voiceless obstruents), [p'] avoided (because it is a weak ejective), [w, j] avoided (for ease of segmenting the speech stream into phones)). All existing words appear in Bricker et al. (1998) and/or *Diccionario Maya Popular* (2004) and were deemed to be in common usage by a native speaker (Santiago Domínguez, from Santa Elena). The Santa Elena word list contains 16 nonce forms and the general word list contains 14 nonce forms. The full word list is provided in Appendix A.

[16] When no other common word could be found, a verb was use if it was a frequent lexical item (as judged by a native speaker) and had an easily identifiable root (see Briceño Chel 2006, Ayres and Pfeiler 1997 for details on Yucatec Maya verbal morphology). All verbs used in this study meet this criteria: *e'es* 'show', *a'al* 'say', *éem* 'descend', *u'ub* 'hear', *maan* 'buy', *beet* 'make, do', *bin* 'go'.

Participants read each word from the word list in isolation and in a random order. Each word was printed on a note card (see (2.1)). For an existing form, the card contained the target word in standard orthography in large print with the Spanish equivalent of the word in smaller print below. The Spanish translation was given because most participants had more reading experience with Spanish and in order to prevent orthographic misunderstandings. For example, a reader should be less likely to mistake *k'a'an* (with a GLOTTALIZED vowel) for *k'áan* (with HIGH TONE) if *k'a'an* is printed alongside *fuerte* 'strong' (as opposed to *k'áan* 'hammock'). With nonce forms, each card contained the target word in standard orthography in large print with an existing word containing the same vowel shape in smaller print below. This format was again designed to prevent misinterpretation.

(2.1) Examples of stimuli used in production study 1

a. existing form	b. nonce form
púuts'	**a'a'**
aguja	como ***a'al*** 'hablar'

There was a brief training and practice session before recording began. The participant reviewed the correlation between orthography and vowel shape and then read a few test note cards, containing both existing and nonce forms. The recording session was divided into two sections: existing forms were presented first, followed by a break during which the participant was asked if they had any questions, and then nonce forms were presented. Throughout the session, participants were encouraged to ask questions if necessary and to repeat the word as many times as they wished if they were not happy with the original pronunciation.[17] In all, participation took about 20 minutes, and participants were compensated for their time.

Recordings were made with an IBM laptop (running PRAAT (Boersma and Weenink 2006)) and a head-mounted microphone (RadioShack product #33-3012) at a sampling frequency of 44.1 kHz. All measurements were extracted from the recordings using PRAAT. For each target word, the boundaries of the vowel were demarcated so that 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 (see §2.1.3 for more detailed information).

2.1.2 Participants and Procedures for Production Study 2 (Frame Sentences)

As discussed in §1.2.4, Gussenhoven and Teeuw (2008) document some of the intonational effects that influence the pitch contours of the four vowel shapes of Yucatec Maya. They show that, in HIGH TONE vowels, a pitch peak occurs early in words in phrase-final position and late in words in non-phrase-final position, but that the pitch contour of GLOTTALIZED vowels is not affected by phrase position. They also show that glottalization is more likely to occur in phrase-final position than in non-phrase-final position. Finally, they claim that the high tone marker of both HIGH TONE and GLOTTALIZED vowels triggers downstep in a following HIGH TONE or GLOTTALIZED vowel. Unfortunately, their data comes from only four native speakers (all male) who live in the San Francisco area and not in their native communities. Because the results of production study 1 (see §§2.4 and 2.5) indicate a large degree of idiolectal and dialectal variation, I felt it was appropriate to conduct a second production study with a larger number of participants who live in various parts of Yucatan,

[17] Measurements were only taken from one production of each form. If the participant repeated the word, measurements were taken from the last production.

Mexico in an effort to provide further documentation of any phrase-position effects on vowel shape. As we will see in the results sections, the current study presents evidence that is in agreement with Gussenhoven and Teeuw's claims about the positional effects on pitch contours but is in disagreement with their claims about downstep.

Production study 2 took place in June and July 2008. Twenty-seven speakers from Mérida, Santa Elena, Sisbicchén, Xocén, and Yax Che, Yucatan were recorded (see Table 2.2). All participants are fluent in Spanish. All data from one speaker from Santa Elena was rejected because this speaker did not produce the requested sentences. This speaker is excluded from the table below and only 26 speakers are reported on here. Some of the speakers from Mérida and Santa Elena had participated in production study 1 in the previous year; all but four of the speakers also participated in the perception study (see Chapter 3).

Table 2.2: Participant data (production study 2)

location	gender	ages	notes
Santa Elena	10 females	21, 21, 21, 23, 26, 31, 34, 38, 60, 65	one fluent in English
	4 males	23, 44, 64, 69	one fluent in English
Mérida	3 males	20, 25, 48	originally from smaller towns on western side of peninsula
Sisbicchén	3 females	29, 31, 36	have lived in Valladolid
	1 male	18	
Xocén	3 males	20, 35, 55	
Yax Che	2 females	21, 28	

(2.2) frame sentences used in production study 2

A. *phrase-initial*	i.	Chake’ tu ya’alaj.				
	ii.	[tʃa.keʔ .tu .jáa̰.lah]				
	iii.	Chak-e’	t-uy		a’al-aj	
	iv.	red-*top*	*comp.asp-3rd.erg*		say-*comp.trans*	
	v.	*Red (is what) s/he said.*				
	vi.	*Rojo (es lo que) él(la) dijo.*				
B. *phrase-medial*	i.	Yaan u ya’alik chak bejla’e’.				
	ii.	[jàa.n u .jáa̰.lik .tʃak .ɓeh.laʔ.eʔ]				
	iii.	Yaan	uy	a’al-ik	chak	bejla’e’.
	iv.	*compuls.asp*	*3rd.erg*	say-*trans*	red	today
	v.	*S/he has to say red today.*				
	vi.	*Él(la) tiene que decir rojo hoy.*				
C. *phrase-final (post-H)*	i.	Táant u ya’alik chake’.				
	ii.	[táan.t u .jáa̰.lik .tʃa.keʔ]				
	iii.	Táant	uy	a’al-ik	chak-e’	
	iv.	*immed.asp*	*3rd.erg*	say-*trans*	red-*part*	
	v.	*S/he just said red.*				
	vi.	*Él(la) acabo de decir rojo.*				
D. *phrase-final*	i.	Tu ya’alaj chak.				
	ii.	[tu .jáa̰.lah .tʃak]				
	iii.	T-uy		a’al-aj		chak
	iv.	*comp.asp-3rd.erg*		say-*comp.trans*		red
	v.	*S/he said red.*				
	vi.	*Él(la) dijo rojo.*				

line (i): complete frame sentence (in standard orthography) with an example target word (*chak*) (as printed on the stimuli)

line (ii) phonemic transcription of sentence (periods denote syllable boundaries, spaces denote phonological words)

line (iii): morphological composition of frame sentence (word-internal morpheme boundaries denoted by dashes)

line (iv): morpheme by morpheme gloss (content words are translated in regular font, grammatical morphemes are defined in italics)

abbreviations used in line (iv): *top* = topicalizer; *comp.asp* = completive aspect; *3rd.erg* = 3rd person ergative pronoun; *comp.trans* = completive aspect, transitive status; *incomp.trans* = transitive status (used with any non-subjunctive non-completive aspect); *compuls.asp* = compulsive aspect; *immed.asp* = immediate (incompletive) aspect; *part* = particle (marks the end of the phrase under the scope of the aspect marker *táant*)

line (v): English translation

line (vi): Spanish translation (as shown to the participant during the training session).

In this study, participants read 144 sentences. Four different frame sentences with 36 different target words were used. The target word occurs in three different positions with respect to the phrase: phrase-initial, phrase-medial, and phrase-final. Additionally, two phrase-final sentences are distinguished by the presence vs. absence of a preceding HIGH TONE vowel. The sentences are shown in (2.2) and all target words are given in Appendix A. All target words were of the form C_1VC_2, where all vowel shapes were used and consonants were voiceless obstruents, ejectives, or voiced sonorants. Each combination of C_1 type, V type, and C_2 type was used (3x4x3) for a total of 36 target words.

All four frame sentences for a single target word were read consecutively, and, as presented to the participant, were preceded by the target word in Yucatec Maya and its Spanish gloss. Four different random orders of sentences were created, such that the order of target words was randomized and the order of frame sentences within each group was randomized. Each participant read each sentence once, following one of the four random orders. Example stimuli are given in (2.3).

(2.3) example stimuli used for production study 2

púuts' *aguja*	**t'uut'** *loro*
Táant u ya'alik púuts'e'.	Táant u ya'alik t'uut'e'.
Púuts'e' tu ya'alaj.	Yaan u ya'alik t'uut' bejla'e'.
Yaan u ya'alik púuts' bejla'e'.	Tu ya'alaj t'uut'.
Tu ya'alaj púuts'.	T'uut'e' tu ya'alaj.

Participants went through a brief training session, where they read practice stimuli. The practice stimuli contained the Spanish translations of the frame sentences. These translations are not included in the stimuli used during the recording session because they were unnecessary once the participant was familiar with the frame sentences. As shown above, the Spanish translation of each new target word was provided to prevent orthographic misunderstandings.

Participants were instructed to request a break at any time and to ask questions if they were not familiar with a target word during the recording session. The length of participation varied – those who regularly read Yucatec Maya completed the task in about 20 minutes, while other participants took up to an hour. All participants were compensated equally.

The same equipment was used as in production study 1. Recordings were made with an IBM laptop and a head-mounted microphone (RadioShack product #33-3012) at a sampling frequency of 44.1 kHz, and all measurements were extracted from the recordings using PRAAT. For each target word, the boundaries of the vowel were demarcated so that 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 (see §2.1.3).

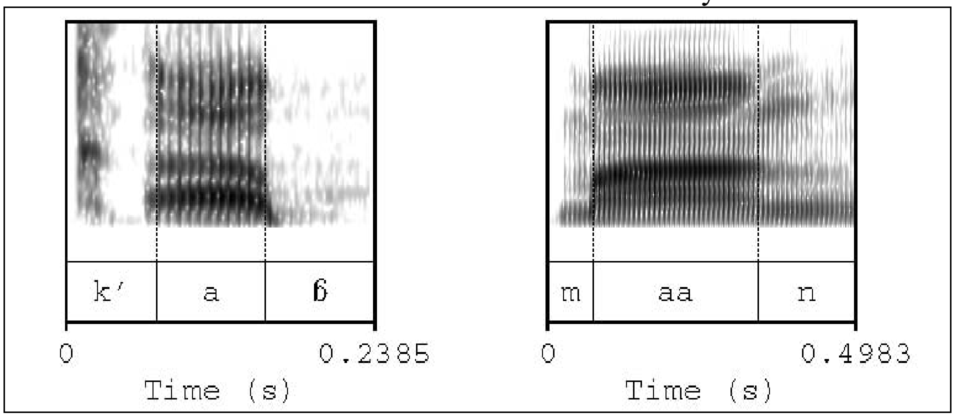
Most of the recordings made for production study 2 are available online at the Archive of the Indigenous Languages of Latin America, hosted by the University of Texas (http://www.ailla.utexas.org). Only those recordings for which the speaker gave permission to use are deposited in the database.

2.1.3 Extraction of Data from Both Production Studies

Data was collected with regard to vowel duration, pitch values produced at various points during vowel production, and type of glottalization produced during vowel production. Vowel duration is calculated by demarcating the nucleus of the target word, based on observance of the spectrogram. The boundaries of the vowel were determined by F_2 onset and offset, as shown in Fig. 2.1.

Figure 2.1: Demarcation of vowel boundaries

a. *k'ab* 'arm' b. *maan* 'buy'



The extraction of pitch measurements and the coding of glottalization type require more discussion, as some of the methodology used is new. I provide detailed explanation of the methodology for both production experiments with respect to measuring pitch in §2.1.3.2 and coding glottalization in §2.1.3.1.

2.1.3.1 Coding Glottalization

It has long been thought that the production of creaky voice results from tight adduction of the vocal folds, which remain just loose enough for voicing (Ladefoged 1971, Laver 1980, Ní Chasaide and Gobl 1995). This places creaky voice near one end of a "glottal stricture continuum":

(2.4) [open] voiceless – breathy – modal – creaky – glottal closure [closed]
(Gordon and Ladefoged 2001)

Edmondson and Esling (2006) argue, through the use of laryngoscopic imagery, that this continuum is too simplified. They identify six "valves of the throat" that are controlled by the speaker to produce various phonation types in the world's languages, and they show that the production of creaky voice is controlled by Valves 1 and 3. Valve 1 refers to the adduction and abduction of the vocal folds, and hence the functioning of this valve in the production of creaky voice is as described in the previous literature – the vocal folds are adducted but not so tightly as to prohibit vibration. Additionally, creaky voice is produced with a constricted Valve 3, which refers to the "sphincteric compression of the arytenoids and aryepiglottic folds forwards and upwards by means of the thyroarytenoid muscle complex" (Table I, p. 159). Creaky voice is distinguished from harsh voice due the latter's use of Valve 2, where the vocal fold vibration is partially covered by the ventricular folds.

Gordon and Ladefoged describe the how creaky voice is identifiable from waveforms:

> "…the creaky phonation is characterized by irregularly spaced pitch periods and decreased acoustic intensity relative to modal phonation. Furthermore, there are fewer pitch periods per second in the creaky token than in the modal one …, indicating a lowered fundamental frequency…" (Gordon and Ladefoged 2001: 387)

They then explicitly identify six acoustic properties that characterize creaky voice as different from modal voice: periodicity, intensity, fundamental frequency, formant frequencies, duration, and spectral tilt. Creaky phonation (as compared to modal phonation) is associated with aperiodic glottal pulses, lower intensity, lower fundamental frequency, higher formant frequencies (at least for the first formant), and (less consistently) longer duration. Spectral tilt is a measure of how rapidly intensity drops as spectral frequency increases. This can be measured by comparing the fundamental frequency (the first harmonic) to the second harmonic – in creaky phonation the intensity of the second harmonic is often equal to or greater than the first harmonic.

In order to categorize the phonation type of each vowel produced in the production studies, the spectrogram and waveform of the target word were observed for the following properties: periodicity, intensity, and fundamental frequency.[18] At the time of coding glottalization type, the stimuli were unlabeled with respect to the target word that each represented.

Fig. 2.2 shows a production of *ti'i'* 'there' where the vowel is produced with modal voice interrupted by creaky voice, and all of the canonical properties of creaky voice are visible. In the central portion of the vowel, the glottal pulses are irregularly and widely spaced (indicative of aperiodicity and low fundamental frequency, respectively) and produced with lower intensity as compared to modal voiced portions on either side.

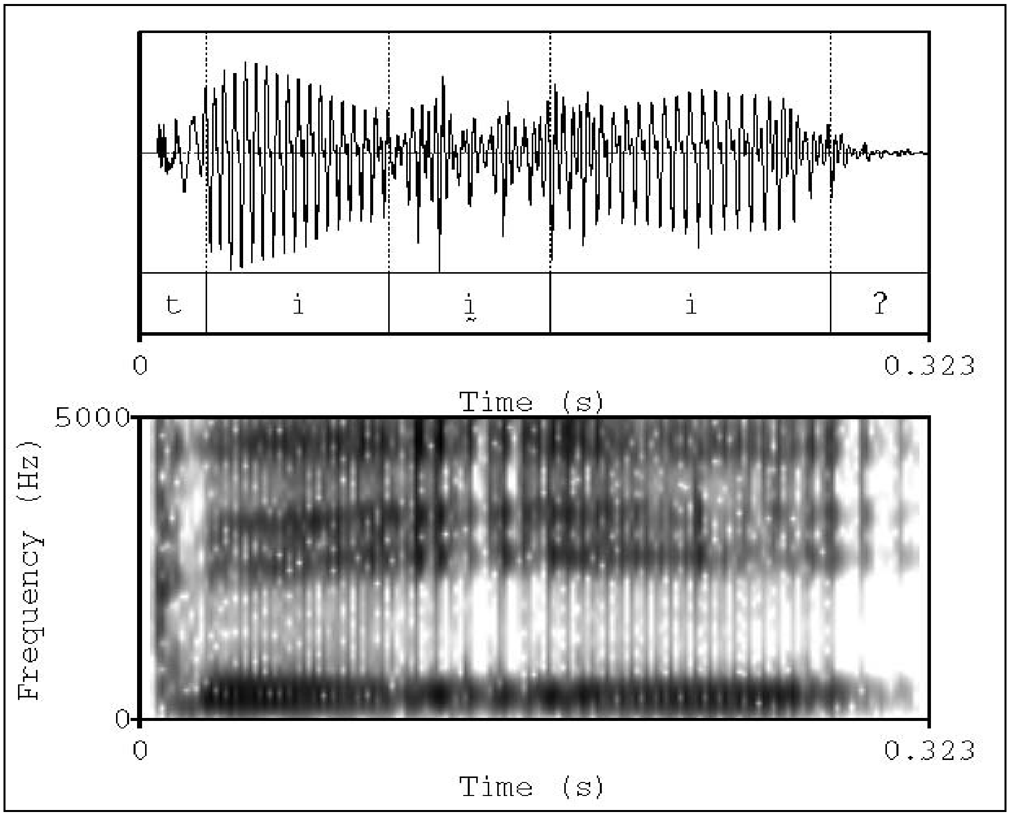
The glottalized portion of this waveform also shows irregularity in the amplitude of consecutive glottal pulses, possibly due to diplophonia,[19] whereas in modal voice, the intensity peaks during successive glottal pulses remain fairly constant or gradually increase or decline. Such erratic jumps and dips in amplitude occurred quite commonly during creaky phonation, though this was never the only indicator of creak; if present it was also coupled with at least an overall weakening of intensity.

[18] Spectral tilt, formant frequency, and duration were not used because these measurements are all correlated with other factors, are less consistently correlated with phonation type, and are not readily observable from waveforms alone. However, it will be shown in §§2.4.1 and 2.5.1 GLOTTALIZED vowels are longer than LOW TONE vowels (but not HIGH TONE vowels) in Santa Elena, and GLOTTALIZED vowels are longer than both LOW TONE and HIGH TONE vowels in Mérida

[19] Diplophonia is the simultaneous production of two distinct tones. While most commonly associated with laryngeal pathology, in rare cases it is intentionally produced due to variation in the tension of each vocal fold (Ward et al. 1969). Redi and Shattuck-Hufnagel (2001: 414) use the visual cues of diplophonia, which they define as "regular alternation in shape, duration, or amplitude of glottal periods", to identify creaky voice. I should note that in my own observations, the alternations between the amplitudes of consecutive glottal periods were often sporadic (and not "regular"), as shown in Fig. 2.2.

Figure 2.2: Common visible characteristics of creaky voice

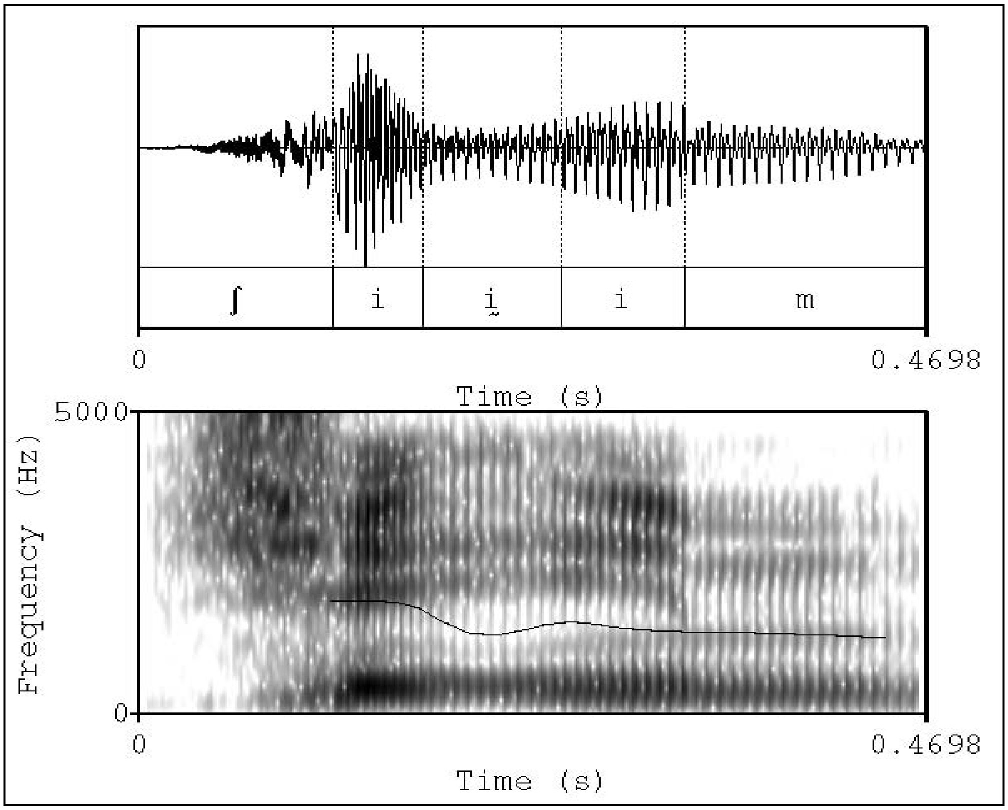
The characteristics of aperiodicity, lower intensity, and widely spaced glottal pulses are seen in this production of ti'i' /tḭiʔ/ *'there' (spoken by a female from Mérida.)*



The classification of this token as having a vowel produced with creaky voice that interrupts modal voice is thus fairly straightforward by using the visible indications of periodicity, intensity, and fundamental frequency. Not all tokens display such nice distinctions between modal and creaky voice, however. In my investigation, I found that intensity was the most consistent visual cue present when the stimulus sounded (to my ear) like it was produced with creaky voice. Additionally, irregularity of the glottal pulses was the least useful characteristic; many productions of GLOTTALIZED vowels sounded creaky and were accompanied by regularly spaced glottal pulses produced with lower intensity and (usually) lower fundamental frequency. One such example is given in Fig. 2.3.

Figure 2.3: Creaky voice as indicated by lower intensity

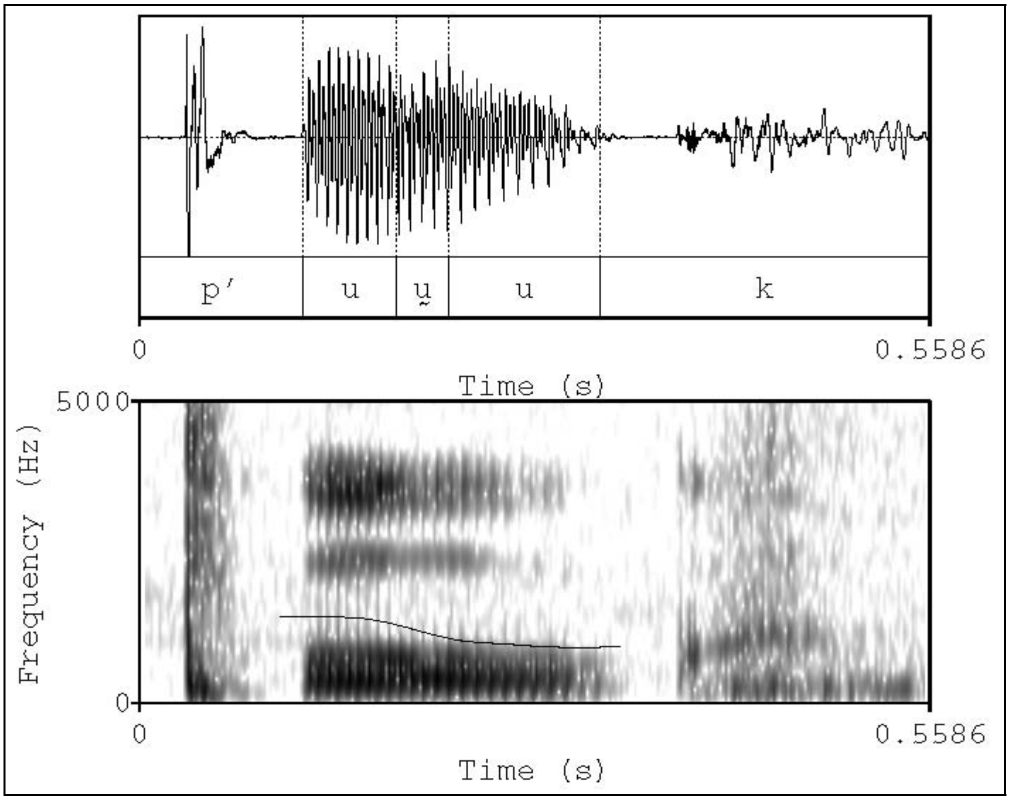
The middle portion of this vowel is produced with a lower intensity and fundamental frequency, but it does not show irregularity in the spacing or amplitude of the glottal pulses. The spectrogram is marked with an uninterrupted pitch contour, indicating regular pitch periods. A male from Mérida produced this token of xi'im /ʃḭim/ *'corn'.*



Tokens such as those presented in Fig. 2.3 – where some portion of the vowel is produced with a dramatic decrease in intensity even though periodic glottal pulses occur throughout production – were classified as being produced with creaky voice. There were many tokens that were even harder to classify in that, while they gave an auditory impression of creak, the only visible cue was a very brief dip in the amplitude of the waveform, as shown in Fig. 2.4. Tokens such as the one below were classified as having "weak glottalization".

Figure 2.4: "Weak glottalization" as indicated by a brief dip in intensity

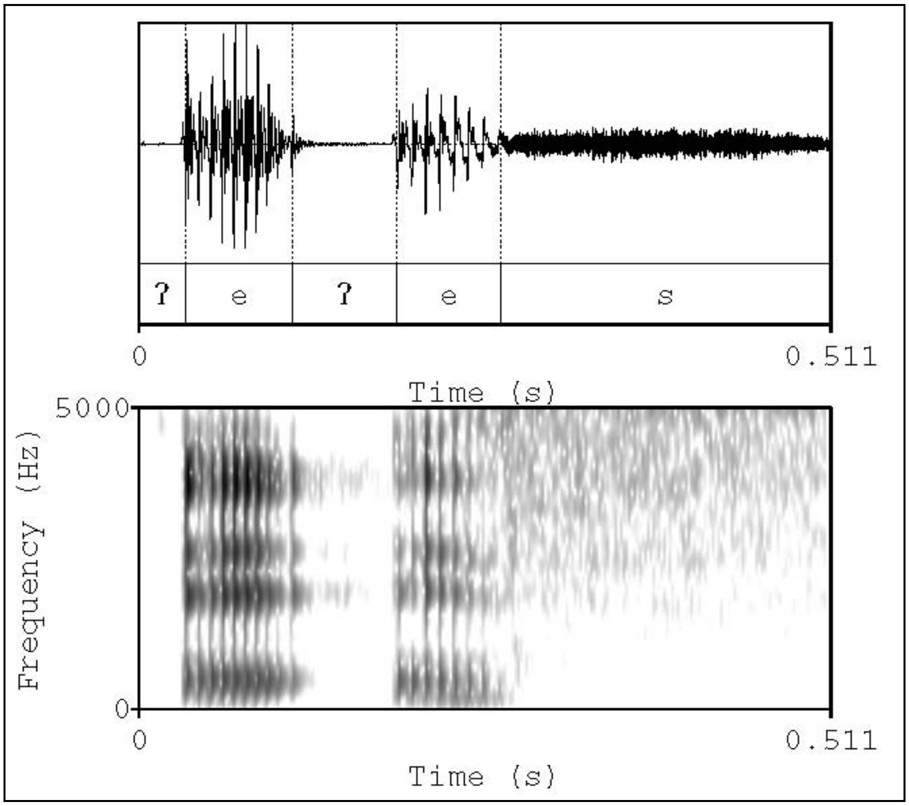
There is a brief period of lower intensity during the middle portion of this vowel, and this token gives an auditory impression of creaky voice. This token represents the category of "weak glottalization", where there are not enough clear indicators of creaky voice or of consistent modal voice. This is a production of p'u'uk /p'úṵk/ *'cheek' as spoken by a male from Mérida.*



The above figures have illustrated the visual cues used to identify creaky voice – lower intensity (the most consistently present cue), lower fundamental frequency, and aperiodicity – and those used to identify the "weak glottalization" category, where the sporadic drop in amplitude of only a few pitch periods indicates some departure from modal voice. In Fig. 2.5, we see the production of a full glottal stop, as indicated by a period of silence (often surrounded by creaky pulses). This production represents the canonical glottal stop in Yucatec Maya GLOTTALIZED vowels: the period of silence is between 70 and 75 ms, and the portion of the vowel that comes after the glottal stop is produced with creaky voice.

Figure 2.5: Full glottal stop as indicated by silence

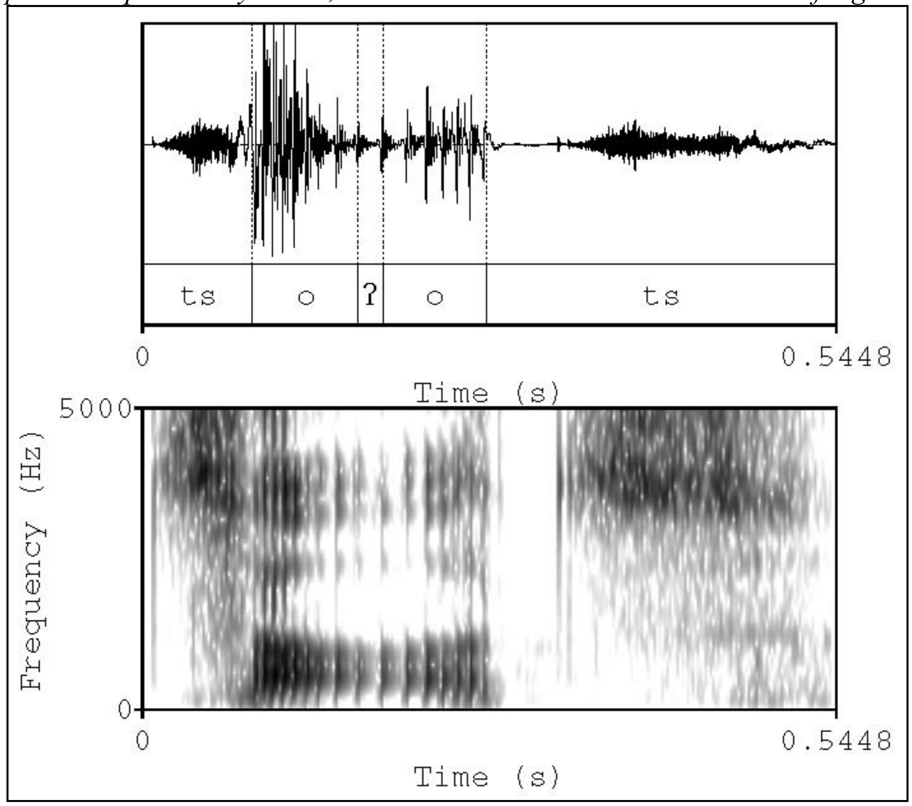
This token – e'es /ʔéḛs/ *'show' – is produced by a male from Mérida.*



Most productions of a full glottal stop resemble the example above, but other tokens do not have such a striking period of silence, and thus were harder to classify as either creaky voice or a glottal stop. In Fig. 2.6, the middle portion of this vowel looks creaky, but the spacing of the pitch periods is quite extreme, even for creaky voice. In the section transcribed with a glottal stop, the pitch periods are 20 ms apart. I used this measurement to indicate a minimal glottal stop – any time two consecutive pitch periods were at least 20 ms apart, I classified the token as being produced with a glottal stop instead of creaky voice.

Figure 2.6: Minimal glottal stop indicated by 20 ms between pitch periods

In this production of tso'ots /tsóo̰ts/ *'hair' (produced by a male from Mérida), there are two pitch periods separated by 20 ms, the smallest distance used as evidence of a glottal stop.*

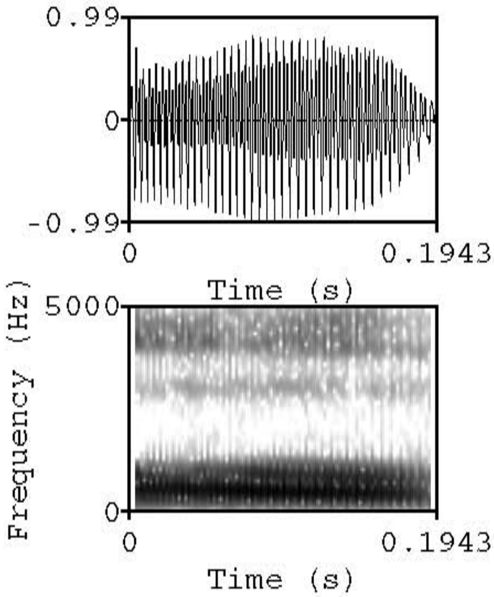


We have now seen examples of the four main categories of glottalization identified in this study: none (modal voice only), weak glottalization, creaky voice, and a full glottal stop. While weak glottalization and a glottal stop were always produced in the middle of the vowel, the production of creaky voice was less consistent. Any production with creaky voice was thus also coded as to which portion of the vowel was creaky: the beginning, the end, the middle, or all. Such coding was based on the indicators of creaky voice and modal voice discussed above.

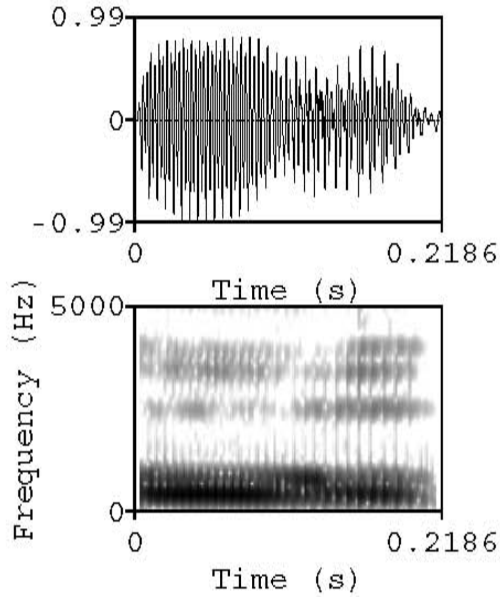
In order to summarize the above discussion on the coding of glottalization, Fig. 2.7 and 2.8 are presented as summaries of the glottalization types identified in this study. These figures only show the vowel, and not the preceding or following consonants of the produced word.

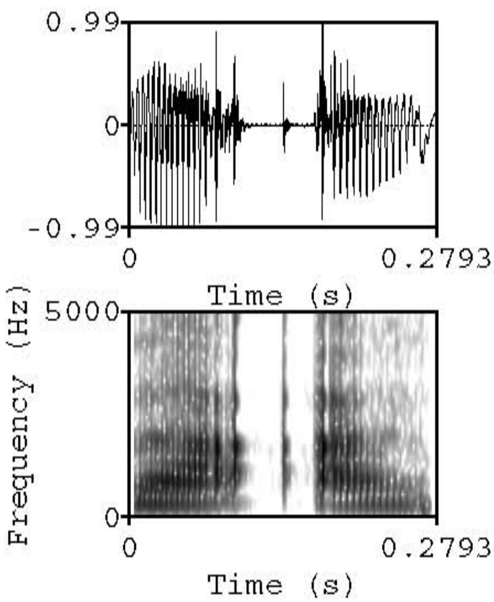
Figure 2.7: Prototypical examples of non-creaky phonation types

a. **modal voice** (from *u'ub* /ʔúṵɓ/ 'to hear', spoken by a female from Mérida)



b. **weak glottalization** (from *u'ub* /ʔúṵɓ/ 'to hear', spoken by a male from Mérida)



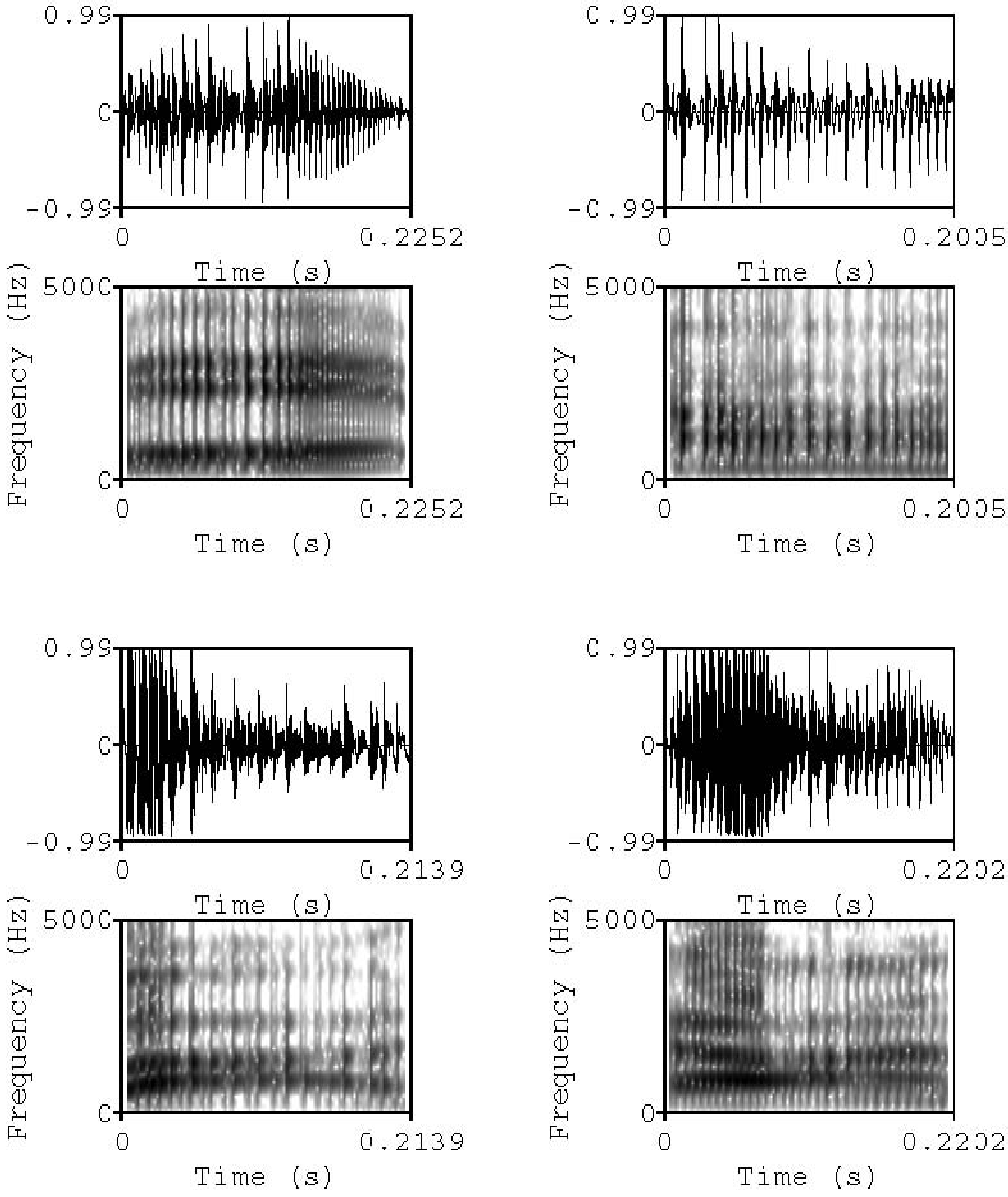


c. **full glottal stop** (from *na'at* /náa̰t/ 'intelligent', spoken by a female from Mérida)

Figure 2.8: Prototypical examples of creaky phonation in different positions

a. **creaky beginning** (from *be'eb* /ɓéḛɓ/ *'type of vine'*, spoken by a female from Sisbicchén)

b. **creaky throughout** (from *a'al* /ʔáa̰l/ 'to say', spoken by a female from Mérida)



c. **creaky end** (from *ba'ax* /ɓáa̰ʃ/ 'what', spoken by a male from Mérida)

d. **creaky middle** (from *k'a'an* /k'áa̰n/ 'strong', spoken by a male from Mérida)

One final note on coding is necessary before we move on to the methodology with regard to pitch. As presented here, the data with respect to glottalization differs from that published in Frazier (2009) due to some changes made in how glottalization was coded. In Frazier (2009), I was rather liberal when coding the presence of creaky voice in that I counted any instance of two or more creaky pulses as creaky voice. It was shown in that paper that the presence of some form of glottalization was correlated with both the pre- and postvocalic consonants. For example, words ending in a glottal stop were more likely to be produced with glottalization. Upon review of the data, it is clear that many of the vowels coded as "creaky end" that were followed by a glottal stop had only a couple of creaky pulses at the end of the vowel and that this creakiness was conditioned by the glottal stop. Since that fact has already been established, it is more useful for our purposes to consider only glottalization that can be assumed to be an intended property of the vowel and not a byproduct of the vowel's environment. For this reason, the data presented here reflects stricter criteria for the coding of glottalization such that at least 4 creaky pulses were required in order to label a vowel as having creaky voice.

2.1.3.2 Measuring Pitch

As mentioned above, pitch measurements were obtained by using PRAAT to extract pitch values (in Hertz) at 10 ms intervals during the production of the vowel. This results in a list of pitch values for each vowel that contains at least four measurements (for the shortest vowels) and up to over 30 measurements (for the longest vowels). From this list, (up to) seven pitch values were retained and will be referred to throughout this paper. The minimum and maximum pitch values for each vowel were identified, regardless of where (in the duration of the vowel) they were produced, and are used to define the *pitch span* of each vowel. Additionally, five pitch values that occurred at normalized time points during the production of the vowel were identified: the first pitch value (time point 1), the pitch value produced at approximately 25% of vowel duration (time point 2), the pitch value produced approximately in the middle of the vowel (time point 3), the pitch value produced at approximately 75% of vowel duration (time point 4), and the last pitch value (time point 5). If pitch was undefined at time points 2-4, then no pitch measurement is recorded for these time points. If pitch was undefined for time point 1 (or time point 5), but was defined 10 ms later (or earlier), then this measurement is used for time point 1 (or 5); otherwise, time point 1 (or 5) is undefined.[20] These five time points are used to define pitch contours throughout this paper.

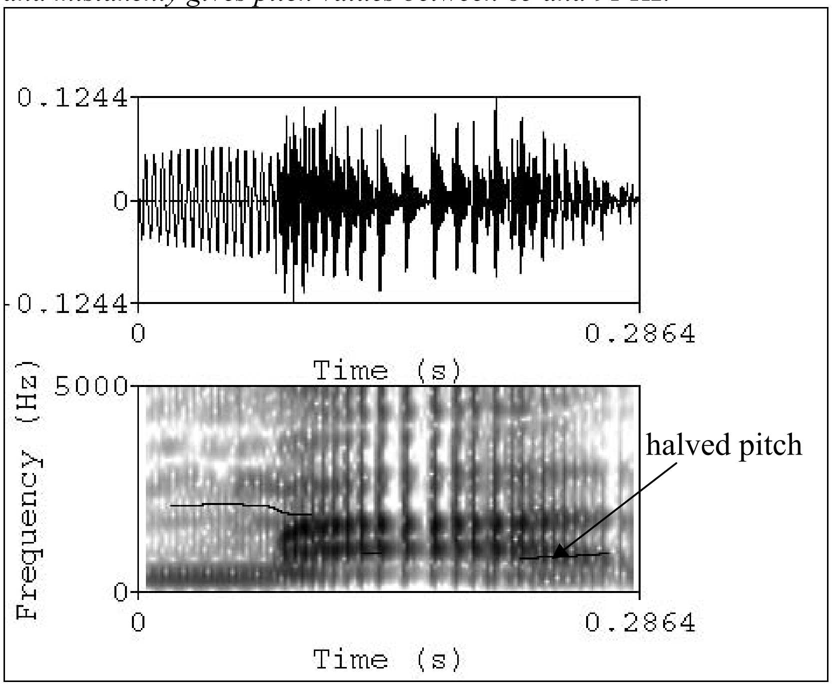
There are many well-known problems associated with measuring pitch. One is that software is not perfect at identifying pitch periods.[21] The most common mistake made by software such as PRAAT is pitch doubling or halving, due to the software treating one pitch period as two or two pitch periods as one. Fig. 2.9 provides an example of pitch halving by PRAAT. Here we see that, in the last portion of the vowel, pitch is very low (the actual values that are reported by PRAAT are between 85 and 91 Hz), but the pitch periods are not widely spaced. It is clear that PRAAT is treating two pitch periods as one, and hence recording pitch at half the value it should be.

[20] Undefined pitch values were treated as missing values and did not result in the removal of any other measurement values with respect to that token.

[21] See Boersma (1993) for a complete discussion of the methodology used by PRAAT to identify and measure fundamental frequency.

Figure 2.9: Example of pitch halving by PRAAT

In this production of ma'ats' [má̰ats'] *'hull (corn)' by a female from Santa Elena (only the [má̰a̰] portion of the word is shown), PRAAT treats two pitch periods as one in the last portion of the vowel and mistakenly gives pitch values between 85 and 91 Hz.*



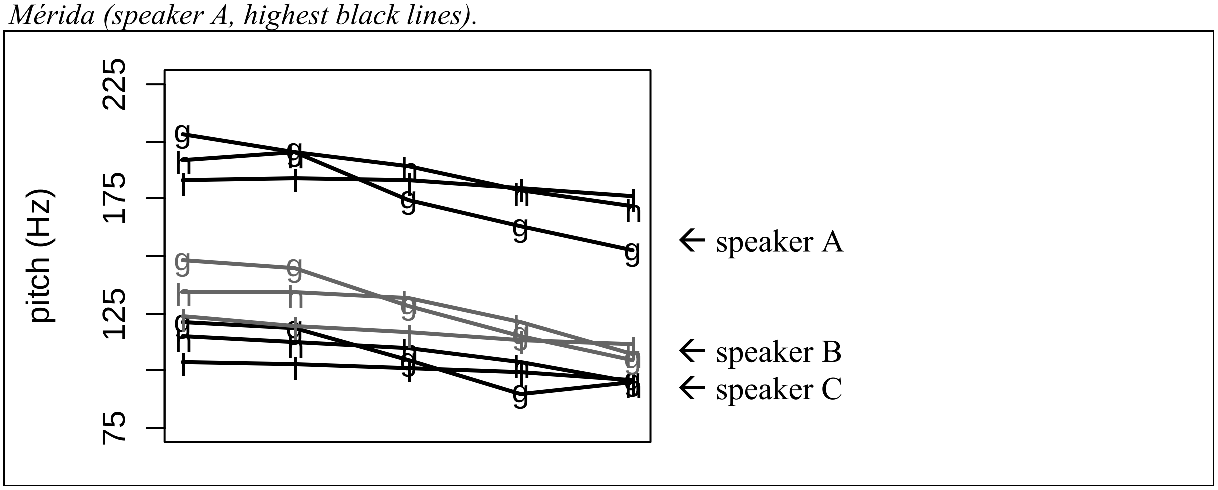
In order to mitigate this problem, a perl script was used to check the data obtained from PRAAT for any instance of one pitch value being more than double or less than half of the previous pitch value. If such a situation occurred, the waveform was checked by hand (by measuring one pitch period) to see if the pitch value recorded by PRAAT was appropriate. If it was, no change was made. If it was not, the original pitch value was doubled, halved, or changed to "undefined" as applicable. The data represented in this dissertation was hence updated in this way, but no other changes were made to the pitch values obtained from PRAAT.

The default settings of PRAAT with regard to pitch measurement were used, such that the acceptable pitch range was 75 – 500 Hz. While a lower ceiling would have been appropriate for all or at least most speakers, I found that using a lower ceiling prompted PRAAT to make more mistakes in terms of halving actual pitch values. For example, some females produced legitimate pitch values over 375 Hz. If the ceiling was set to 375 Hz, PRAAT would record these values at half of the accurate value. I found it best to let PRAAT record values with a high ceiling and then adjust spurious values by hand following the procedure outlined above.

Another problem with measuring pitch comes from attempting to compare pitch values across speakers. It is notoriously difficult to compare utterances by males and females in terms of pitch due to the physiological differences that result in females having a higher natural pitch range than males. The problem is not solely gender-related – any two speakers with different natural pitch ranges will produce considerably different pitch values for the same phonological tone. Fig. 2.10 illustrates this, where we see average pitch contours for the long vowels for three different participants, two males and one female, in production study 1 (isolation). Even though the general shape of the contours for each vowel shape is similar among the three speakers, the pitch values are quite different.

Figure 2.10: Production of pitch by three different speakers

This graph displays the average pitch contours produced in production study 1 for LOW TONE *(*'l'*),* HIGH TONE *(*'h'*), and* GLOTTALIZED *(*'g'*) vowels, as produced by a male from Mérida (speaker C, lowest black lines), another male from Mérida (speaker B, middle gray lines), and a female from Mérida (speaker A, highest black lines).*



One way to abstract away from the differences caused by natural pitch range is to measure a *pitch span*, which is a measurement of the difference between the high and low pitch points of a pitch contour (Ladd 1996), instead of a single pitch value. However, Hertz is still a problematic measurement of pitch spans. Again, the problem is due to differences in natural pitch range. Using the same speakers represented in Fig. 2.10, if we calculate the average difference between the maximum and minimum pitch values produced for HIGH TONE vowels for each speaker, we see that speaker C (with the lowest pitch range) has the smallest average pitch span (22.5 Hz), while speaker B (with the next lowest pitch range) has the next smallest average pitch span (30.8 Hz), and speaker A (with the highest pitch range) has the largest average pitch span (35.2 Hz).

In perception, pitch spans at the low end of the Hertz scale are perceived as more different than pitch spans of equal length at the high end of the Hertz scale. In other words, even though the speakers represented in Fig. 2.10 produced HIGH TONE vowels with different absolute pitch values and with different pitch spans, this does not mean the listener will perceive the pitch span for the female as larger than those for the males. This phenomenon has lead to the development of a variety of psycho-acoustic scales to measure pitch, including semitones, mels, Bark, and ERB-rate. According to Nolan (2003), semitones provide the best measurement for pitch spans in intonational contours, such that differences among speakers are minimized. The formula for conversion of Hertz to semitones is shown in (2.5).[22] Using this formula, if the reference Hz is 100, then the semitone measurement would denote semitones over 100 Hz.

[22] Those who study western music will recognize a semitone as 1/12 of an octave. For example, on a piano there are twelve (black and white) keys for each octave, each successive key representing an increment of a semitone.

(2.5) semitones = $12*\log_2(\text{Hz}/\text{reference Hz})$
where reference Hz is an arbitrary constant

Even with the conversion of Hz to semitones, it is still necessary to measure pitch spans instead of individual pitch values in order to make valid comparisons across speakers. If we convert the average pitch value (in Hz) produced at the beginning of HIGH TONE vowels by each participant shown in Fig. 2.10 to semitones over 100 Hz, we get 2.4 for speaker C, 5.2 for speaker B, and 11.2 for speaker C, for an average of 6.3 semitones. This number – 6.3 semitones – is not meaningful in that we cannot use it to predict what speaker D might produce for the initial pitch value of this contour. If we analyze spans instead, using the equation $12*\log_2(\text{max Hz}/\text{min Hz})$, the average pitch span for speaker C is 3.7 semitones, 4.4 semitones for speaker B, and 3.9 semitones for speaker C, for an average of 4.0. This value does have meaning in that it can be used to predict the pitch span that would be produced by other speakers with different natural pitch ranges.

In measuring the pitch of the target words used in production studies 1 and 2, not every vowel shape that was produced has a pitch span that is meaningful. For example, the LOW TONE vowels are produced with fairly steady low pitch. There will of course be minimum and maximum pitch values produced during any pitch contour, but in this case these extreme values cannot be expected to occur in stable positions within the vowel nor can the difference between them be expected to be meaningful. For this reason, I do not measure a traditional pitch span, but instead I use a constant relative to each speaker to scale pitch values from that speaker. This method is based on Pierrehumbert's (1980) work with English intonation. In her dissertation, Pierrehumbert found that peaks in an intonation contour varied relative to the pitch value produced for the final low boundary tone. She calls this value for the low boundary tone a *baseline*, which is unique for each speaker, and measures pitch with the formula: (pitch – baseline)/baseline.

Again, there are differences between Pierrehumbert's target of analysis (intonational contours) and my own (pitch measurement of both level and contour tones), such that I cannot measure a baseline in the same way. I will, however, make use of her identification of speaker-specific baseline. I define the baseline for each speaker as the average pitch value produced at the middle point of LOW TONE vowels (as spoken in a particular context), and I pick this point because it is both low and relatively stable. The baseline is defined for each context. For example, pitch values at the mid point of all LOW TONE vowels spoken in isolation in production study 1 are averaged together to get the baseline for a given speaker in production study 1. I thus measure pitch in production study 1 according to the formula in (2.6). As mentioned above, pitch values are first extracted in Hertz. After calculating the baseline for each speaker, these pitch values can then be transformed into *semitones over the baseline* (s/b). Thus, a pitch value of 3.4 s/b denotes a pitch value that is 3.4 semitones above *that speaker's* baseline. In this way, pitch values from different speakers can be averaged together and those averages will have meaning.

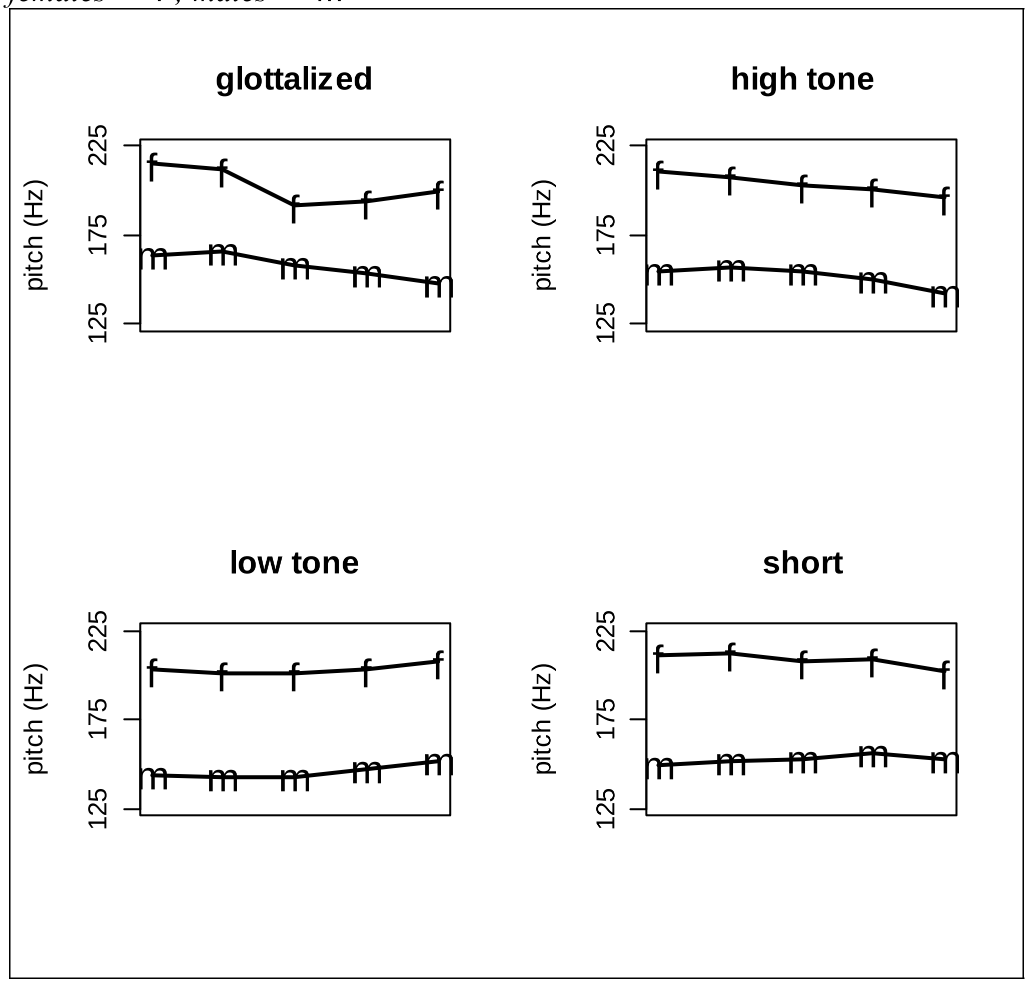
(2.6) equation used in production study 1 for the transformation of Hertz to semitones over the baseline (s/b)

$$s/b = 12*\log_2(Hz/\text{baseline } Hz)$$

where baseline Hz is the average pitch value produced at the mid point of LOW TONE vowels ***for a given speaker***

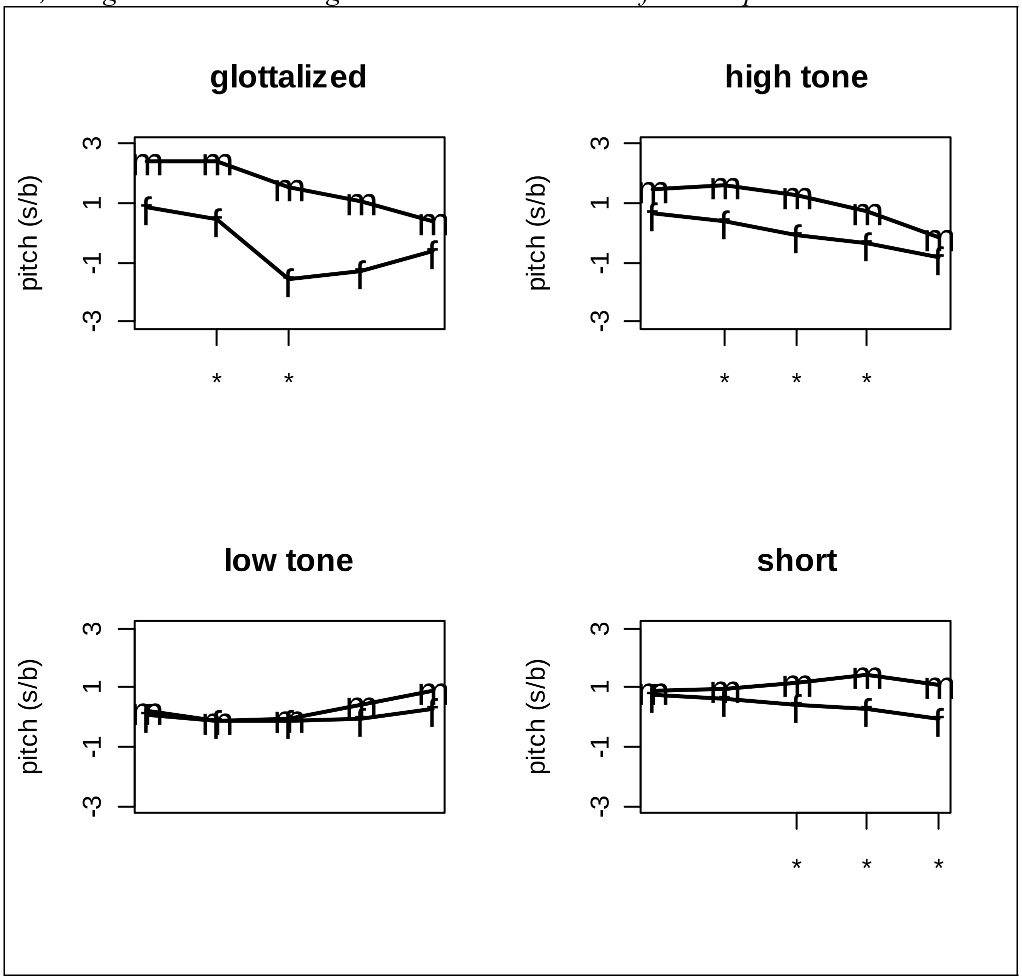
In order to assess the success of this transformation, we will look at the average pitch contours produced by speakers from Santa Elena in production study 1. In Fig. 2.11 we see average pitch contours by gender as measured in Hertz. With the exception of the GLOTTALIZED vowel, the shapes of the contours are about the same for men and women, but the values at each time point are (as expected) very far apart. If we average both genders together, we would get average pitch contours that are not representative of either gender – the average pitch values would be essentially meaningless.

Figure 2.11: Pitch measured in Hertz
average pitch contours for all speakers from Santa Elena (production study 1)
females = 'f'; males = 'm'



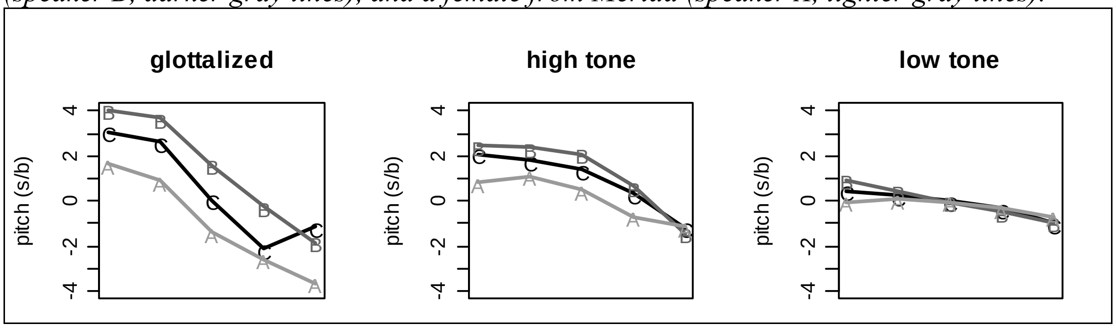
On the other hand, Fig. 2.12 shows the same average pitch contours as measured in semitones over the baseline. In this figure, asterisks mark statistically significant differences between the genders for that time point. While there are some statistically significant differences, the pitch contours for HIGH TONE, LOW TONE, and SHORT vowels are remarkably similar for both genders. The GLOTTALIZED vowels, however, are different not only in the values at each time point but also in terms of the overall shape of the contour for the two genders. This result will be discussed in the next section.

Figure 2.12: Pitch measured in semitones over the baseline
average pitch contours for all speakers from Santa Elena (production study 1, isolation)
females = 'f'; males = 'm'; asterisks denote statistically significant differences between genders (p < .05, using a mixed linear regression model to account for multiple observations within subjects)



One final figure will be useful for documenting the relative success of the semitone transform. Fig. 2.13 shows the average pitch contours in s/b for the same three speakers as were presented in Fig. 2.10. In this new figure, the HIGH TONE and LOW TONE vowels have remarkably similar pitch contours for all speakers, whereas the GLOTTALIZED vowel shows more inter-speaker variation. When we measure pitch in s/b during the production of modal voice (as evidenced by the HIGH TONE, LOW TONE, and SHORT vowels), we can obtain meaningful averages across speakers of both genders. For the non-GLOTTALIZED vowels, it is clear that the semitone transform provides us with data that is more comparable between speakers of different genders than pitch measurements in Hertz.

Figure 2.13: Production of pitch by three different speakers (s/b)
This graph displays the average pitch contours as produced in production study 1 by the same speakers in Fig. 2.10. (a male from Mérida (speaker C, black lines), another male from Mérida (speaker B, darker gray lines), and a female from Mérida (speaker A, lighter gray lines).



For production study 2, the measurement of pitch was done in the same fashion. However, because we are comparing pitch contours in vowels that are produced in different phrase positions, we want to minimize the effects of declination (the gradual lowering of F_0 that happens throughout an utterance). In order to do this, a context-dependent baseline is used for each speaker, such that the baseline is the average pitch value produced for the middle pitch point of LOW TONE vowels *in a particular frame sentence*. This means that there are four different baselines for each speaker, and each baseline is used only for target words produced in the same frame sentence. The equation in (2.6) is still used to transform Hertz into semitones.

2.2 The Interaction of Pitch and Creaky Voice

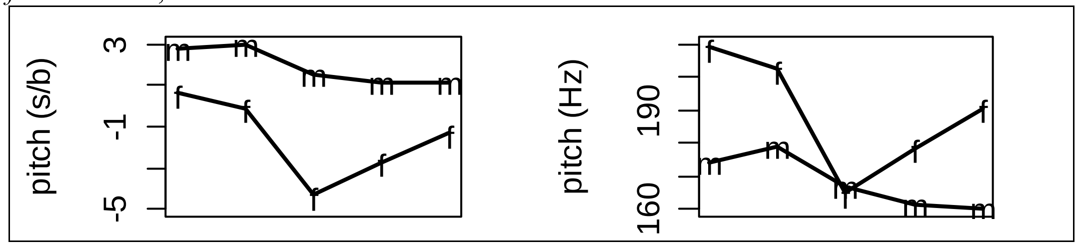
A surprising result was represented in Fig. 2.12, which is that the semitone transform does not work as well with GLOTTALIZED vowels as it does with modal voiced vowels. In this section I explain how this is due to an interaction between pitch and creaky voice. GLOTTALIZED vowels tend to be produced with creaky voice in the middle of the vowel. Creaky voice is, of course, associated with low pitch, as discussed in §2.1.3.1. What the data from Yucatec Maya shows us is that the low pitch (in Hz) associated with creaky voice is constant for men and women, and hence unrelated to the speaker's natural pitch range. The baseline on average is, of course, higher for females than for males, and so productions of creaky voice by males result in pitch that is slightly above their baseline, while productions of creaky voice by females result in pitch that is much lower than their baseline.

This is made explicit in Table 2.3, which provides average pitch values for time points 1-5 for each gender as measured in both Hz and s/b for only those GLOTTALIZED vowels that are produced with some form of glottalization, and Fig. 2.14, which charts those same pitch contours. We see in Fig. 2.14 that when we measure pitch in Hertz, we get the same average values for males and females in the middle of the vowels, where creaky voice is canonically produced. Thus, if we transform this value to semitones over the baseline, we get a negative number for females and a positive one for males. Furthermore, because the pitch associated with creaky voice is so low (relative to the natural pitch range for the average female), it seems that this affects the rest of the pitch contour for females, and this is why the semitone transform is less successful even for the modal voiced portions of the GLOTTALIZED vowel. In other words, females are not producing pitch that is as high (relative to their baseline) before or after creaky voice because this would result in too much of a change in pitch between successive time points.

Table 2.3: Average pitch values for GLOTTALIZED vowels produced with some form of glottalization (Santa Elena only, production study 1)

	Hertz		s/b	
	males	females	males	females
1	174.1	209.0	2.8	0.6
2	179.1	202.2	2.9	-0.1
3	167.0	164.8	1.5	-4.3
4	161.6	178.5	1.1	-2.7
5	160.4	190.1	1.1	-1.3

Figure 2.14: Average pitch contours for GLOTTALIZED vowels produced with some form of glottalization (Santa Elena only, production study 1)
females = 'f'; *males* = 'm'



We can conclude from this discussion that, when creaky voice is involved, measurements of pitch in semitones over the baseline will differ more between males and females than when measurements of pitch come from only modal voiced segments. In the results that are presented in §§2.4 and 2.5, I will address this interaction when discussing GLOTTALIZED vowels, but I will also abstract away from this complication when possible and look at pitch contours where males and females are averaged together. This result will be highly relevant for the theoretical discussion of Chapter 5, and so we will return to it at that time.

2.3 Methodology Summary

In §§2.1 and 2.2 I have presented the methodologies of production studies 1 and 2. We will briefly review these methodologies here.

In production study 1, 24 participants from Santa Elena, Mérida, and Sisbicchén read 100 target words (mostly of the form CVC) in isolation. In production study 2, 26 participants from Santa Elena, Mérida, Sisbicchén, Xocén, and Yax Che read target 36 words of the form CVC, each in four different frame sentences. The target word occurred in phrase-final, -medial, and -initial positions. There were two sentences where the target word was in phrase-final position; one where it occurred after a word with a HIGH TONE vowel and one where it did not.

For each target word, vowel length was measured, pitch at five normalized time points was measured, maximum and minimum pitch values (during vowel production) were obtained in order to calculate a pitch span, and glottalization type was coded. Four main categories of glottalization were used: none (modal voice produced through vowel production), weak glottalization (denoted by brief dip in the intensity of the wave form), creaky voice (denoted most often by a weakening of intensity and less reliably by aperiodicity, low F_0 and sporadic intensity of glottal pulses), and a full glottal stop (denoted by at least 20 ms of silence). The position of creaky voice with respect to vowel boundaries was also coded.

The pitch values obtained from PRAAT are measured in Hertz. These values are then transformed into semitones over the baseline (s/b) where the baseline is the average pitch value of all LOW TONE vowels at time point 3 *produced by a given speaker in a given context*. This means that the baseline used in the semitone transform is specific to each speaker and each context (isolation or one of the four frame sentences). This transform is designed to allow for valid cross-speaker comparisons of pitch measurements.

While the semitone transform is quite successful in comparing measurements from multiple speakers when modal voice is produced, it is not successful when creaky voice is produced. This is because creaky voice is associated with a constant pitch value for both males and females. This constant is slightly above the baseline for males and far below the baseline for females. There is thus an interaction among pitch, creaky voice, and gender that cannot be ignored. This interaction will be addressed when necessary in the results sections of this chapter and in the theoretical discussion of Chapter 5.

2.4 Results for Participants from Santa Elena

2.4.1 Vowel Length

In Santa Elena, HIGH TONE, LOW TONE, and GLOTTALIZED vowels are almost twice as long as SHORT vowels, when produced in target words in isolation (Table 2.4), and GLOTTALIZED and HIGH TONE vowels are significantly longer than LOW TONE vowels.

Table 2.4: Vowel length in Santa Elena (production study 1, isolation)

	mean length (ms)	standard deviation
SHORT	119	59
LOW TONE	197	92
HIGH TONE	214	88
GLOTTALIZED	220	93

statistically significant differences:
H, G > L > S ($p \leq .01$ using a mixed linear regression model to account for multiple observations within subjects)[23]

In order to determine how much inter-speaker variation there is in the production of vowel length, Table 2.5 provides the average vowel lengths for each vowel shape as produced by each participant from Santa Elena. While there is variation in how long the vowels are, the long vowels are always clearly distinguished from the SHORT vowels. Furthermore, if there are any statistically significant differences among long vowels, either the GLOTTALIZED vowel is longer than the LOW TONE vowel or both the GLOTTALIZED and HIGH TONE vowels are longer than the LOW TONE vowel. This data essentially shows that speakers vary in speaking rate (as is to be expected) but that vowel length is clearly contrasted no matter what the speaking rate is and that if one long vowel is the longest it is the GLOTTALIZED vowel.

[23] Throughout this dissertation, I summarize the results of statistical analysis with a 'greater then' sign, e.g. $x > y$, to denote that the mean of x is significantly greater than y. A comma indicates no significant difference. In such summaries it is always the case that all pairwise comparisons were tested.

Table 2.5: Vowel length for each participant from Santa Elena (production study 1, isolation)
Average vowels length (ms) for each participant from Santa Elena by vowel shape. All differences between a long vowel and the SHORT vowel are statistically significant; If any differences among long vowels are statistically significant, they are noted below the average vowel lengths for that speaker (p < .05, using a linear regression model)

	1	2	3	4	5	6	7	8	9	10	11	12
SHORT	113	122	136	106	75	137	205	142	92	103	120	83
LOW TONE	175	172	224	212	142	279	334	205	164	150	158	148
HIGH TONE	210	190	254	239	165	236	336	246	181	161	163	187
GLOTTALIZED	199	205	247	253	182	258	326	258	178	175	172	186
				G>L	G>L			G>L				G,H>L

The long vowels are also almost twice as long as the SHORT vowel when produced in frame sentences (Table 2.6). Here there are no significant differences among the long vowels, and there is only a slight effect of frame sentence: GLOTTALIZED vowels are longer in sentence D than in sentence A and HIGH TONE and LOW TONE vowels are longer in sentence D than in sentence B. There are no statistically significant differences for the SHORT vowels by frame sentence. These results indicate slight phrase-final lengthening for the long vowels only.

Table 2.6: Vowel length in Santa Elena (production study 2, frame sentences)

	frame sentence			
	A	B	C	D
	phrase-init.	*phrase-med.*	*phrase-final (post-H)*	*phrase-final*
	__e’ tu ya’alaj.	Yaan u ya’alik __ bejla’e’.	Táant u ya’alik __e’.	Tu ya’alaj __.
SHORT	113	117	112	108
LOW TONE	179	169	176	187
HIGH TONE	179	170	177	186
GLOTTALIZED	172	176	176	186

It is clear from the above tables that length is a distinctive feature for speakers from Santa Elena that distinguishes all long vowels (LOW TONE, HIGH TONE, GLOTTALIZED) from the SHORT vowel.

2.4.2 Glottalization

If we look at all four vowel shapes, we see that glottalization is marginally present in the non-GLOTTALIZED vowels: 10% of HIGH TONE vowels and 4% of LOW TONE and SHORT vowels are produced with some form of glottalization (weak glottalization, creaky voice, or a full glottal stop) in production study 1 (isolation). Of the non-GLOTTALIZED vowels produced with glottalization, many of them are nonce forms. If we look at just the existing forms, the percentage of tokens produced with glottalization changes to 3% for HIGH TONE and SHORT vowels and 0% for LOW TONE vowels. Furthermore, the results of production study 2 (frame sentences), where only existing forms were used, show that 2% of all non-GLOTTALIZED vowels are produced with some form of glottalization. Interestingly, in production study 1, glottalization in GLOTTALIZED vowels is not affected by whether the word is an existing form or a nonce form. This means that the speaker is more likely to produce glottalization when producing an unknown form, but only if that form is supposed to be produced with modal voice. In the rest of this section, I discuss glottalization in the GLOTTALIZED vowels only and ignore any differences between existing and nonce forms.

The most common types of creaky voice found in this study match those noted in Avelino et al. (2007): "creaky middle" and "creaky end". A glottal stop was produced less often than both of these types of creaky voice. The distribution of glottalization types for GLOTTALIZED vowels is summarized in Table 2.7. A small percentage of vowels are produced with creaky voice throughout the duration of the vowel. While no tokens from Santa Elena were produced with creaky voice only at the beginning of the vowel, I include the "creaky beginning" category because it will appear in §2.5.2 when I discuss data from the other participants. Many of the GLOTTALIZED vowels fall in the "weak glottalization" category. This means that if we were to use even stricter criteria for the classification of a vowel as being produced with glottalization, such that these vowels would not merit the label, only 37% of GLOTTALIZED vowels would be coded as having glottalization.

Table 2.7: Distribution of glottalization types for GLOTTALIZED vowels in Santa Elena (production study 1, isolation)

full glottal stop	6.7%
creaky middle	20.3%
creaky end	8.0%
creaky throughout	1.7%
creaky beginning	0.0%
weak glottalization	16.7%
none (modal voice)	46.7%

The distribution of glottalization types is affected by gender, as shown in Table 2.8. Here we see that the females are more likely to produce glottalization and that, when glottalization is produced, females are most likely to produce creaky voice in the middle of the vowel, whereas males are most likely to produce weak glottalization.

Table 2.8: Distribution of glottalization types for GLOTTALIZED vowels by gender in Santa Elena (production study 1, isolation)

	males	females
full glottal stop	2.4%	9.7%
creaky middle	10.4%	27.4%
creaky end	5.6%	9.7%
creaky throughout	3.2%	0.6%
weak glottalization	22.4%	12.6%
none (modal voice)	56.0%	40.0%
Rao-Scott $\chi^2(5) = 17.5$,[24] p = .004; null hypothesis of equal distributions between genders		

We should be cautious about interpreting the data with respect to gender because there is significant inter-speaker variation in the production of glottalization as shown in Table 2.9 (which uses simplified categories of glottalization). For example, looking at just the females, participant 3 never produces a glottal stop, while participant 4 produces a glottal stop a fifth of the time. Given this much variation among participants, it is unlikely that we

[24] The Rao-Scott χ^2 is analogous to Pearson's χ^2 and is adjusted for multiple observations within subjects.

can take the gender differences shown in Table 2.8 as representative of the whole population.

Table 2.9: Distribution of glottalization types for GLOTTALIZED vowels by participant (production study 1, isolation)

	females						
type of glottalization	1	2	3	5	8	10	11
glottal stop	4%	4%	0%	20%	16%	8%	16%
creaky voice	52%	32%	8%	52%	20%	56%	44%
weak glottalization	4%	12%	12%	20%	16%	8%	16%
none (modal voice)	40%	52%	80%	8%	48%	28%	24%
	males						
	4	6	7	9	12		
glottal stop	0%	8%	0%	0%	4%		
creaky voice	28%	0%	40%	4%	24%		
weak glottalization	36%	4%	32%	8%	32%		
none (modal voice)	36%	88%	28%	88%	40%		

Table 2.10 presents the distribution of glottalization types for each frame sentence as obtained from production study 2. We see a general decrease in the production of only modal voice and increase in the production of creaky voice or a glottal stop as the target word moves toward the right edge of the sentence. This matches the result found by Gussenhoven and Teeuw (2008).

Table 2.10: Distribution of glottalization types by sentence for GLOTTALIZED vowels in Santa Elena (production study 2, frame sentences)

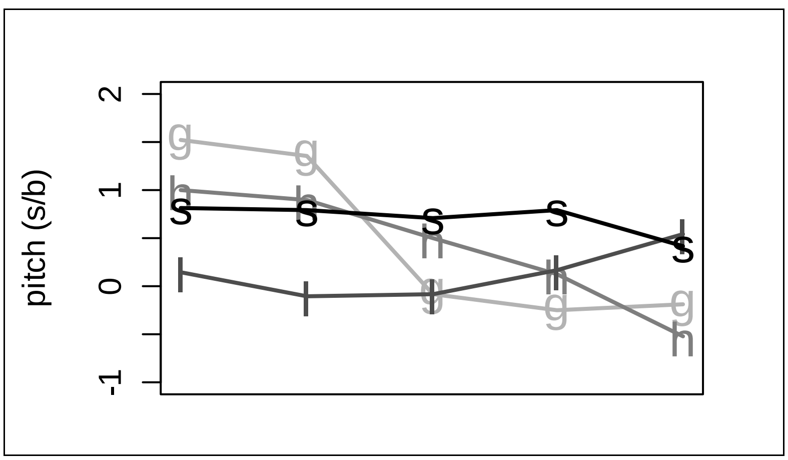
	frame sentence			
	A	B	C	D
	phrase-init.	*phrase-med.*	*phrase-final (post-H)*	*phrase-final*
	__e' tu ya'alaj.	Yaan u ya'alik __ bejla'e'.	Táant u ya'alik __e'.	Tu ya'alaj __.
glottal stop	0.0%	0.0%	0.8%	0.8%
creaky voice	3.9%	13.5%	19.1%	25.5%
weak glottalization	4.0%	7.9%	10.3%	9.5%
none (modal voice)	92.1%	78.6%	69.8%	64.2%

In this section we have seen evidence that glottalization is a prominent feature of the GLOTTALIZED vowel, which distinguishes this vowel shape from the other three. When producing glottalization, there is a general preference for creaky voice or weak glottalization instead of a glottal stop, and if creaky voice is produced it is most likely to occur in the middle of the vowel. The likelihood of producing glottalization increases as the target word moves towards the end of the sentence.

2.4.3 Pitch

In Fig. 2.15 we see the average pitch contours for each vowel shape as produced by all speakers from Santa Elena in production study 1. The statistically significant differences among the pitch values for each vowel shape as produced at each time point are summarized in Table 2.11.[25] In this table, the "greater than" sign denotes that the vowel shape(s) to the left has an average F_0 (at that particular point) that is significantly greater than the average F_0 of the vowel shape(s) to the right. The symbol is used transitively, e.g. for initial pitch, GLOTTALIZED vowels have a significantly higher F_0 than all three other vowel types.

Figure 2.15: Average pitch contours for Santa Elena (production study 1, isolation)
GLOTTALIZED = 'g'; HIGH TONE = 'h'; LOW TONE = 'l'; SHORT = 's'



[25] Because most of the graphs in this chapter display multiple lines, I found that using error bars in the graphs created a rather chaotic display. For this reason, the results of statistical testing are generally presented in tables. The reader should be cautious when drawing inferences from the graphs alone because, when applicable, all test statistics in this dissertation are adjusted for multiple observations within subjects. This can result a significant difference between two points that look close together and/or a nonsignificant difference between two points that look far apart.

Table 2.11: Statistical analysis of pitch among vowel shapes in Santa Elena (production study 1, isolation)

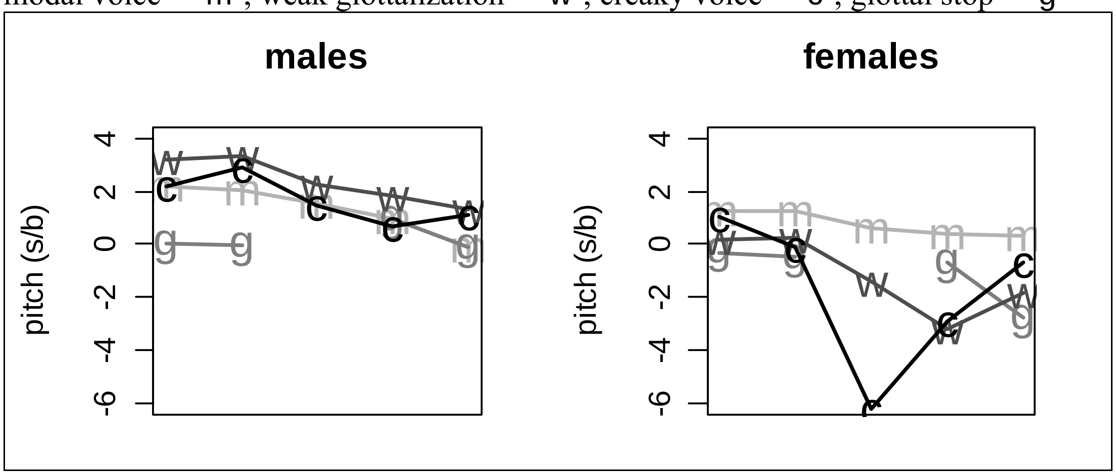
	statistically significant differences among vowel shapes (p <.05, using a mixed linear regression model to account for multiple observations within subjects)
1	GLOTTALIZED > HIGH TONE, SHORT > LOW TONE
2	GLOTTALIZED> HIGH TONE, SHORT > LOW TONE
3	HIGH TONE, SHORT > GLOTTALIZED, LOW TONE
4	SHORT > HIGH TONE, LOW TONE, GLOTTALIZED
5	SHORT, LOW TONE > GLOTTALIZED, HIGH TONE

To summarize Fig. 2.15 and Table 2.11, GLOTTALIZED vowels start with extremely high pitch which then drops sharply; HIGH TONE and SHORT vowels start with high pitch, but the HIGH TONE vowels have a falling contour, whereas the SHORT vowels do not; and LOW TONE vowels start with low pitch and have a slight rising contour. Based on Fig. 2.15 alone, there seems to be an argument for labeling GLOTTALIZED vowels as having high tone and HIGH TONE (and possibly SHORT) vowels as having mid tone. However, as discussed in §1.2.4, both HIGH TONE and GLOTTALIZED vowels have been described in the literature as having high tone and SHORT vowels have been described as being "neutral" with respect to tone. As we look at the data in the rest of this section and in §2.5.3, we will see that the SHORT vowel does not always have pitch that is quite so high and that the pitch of GLOTTALIZED vowels is not always higher than HIGH TONE vowels. For these reasons, I prefer to remain consistent with the literature in terms of the labeling of tones: there is clear evidence in Fig. 2.15 that HIGH TONE and GLOTTALIZED vowels are differentiated from LOW TONE vowels in that the first two are marked for high tone and the latter for low tone.

As discussed in §2.2, there is an interaction between pitch and creaky voice that causes the pitch contours for the GLOTTALIZED vowels to differ by gender. In order to develop a more thorough picture of exactly how pitch is affected by glottalization, Fig. 2.16 shows the average pitch contours of GLOTTALIZED vowels by gender and glottalization type.

Here we see that those GLOTTALIZED vowels produced with modal voice have remarkably similar contours for both genders (only the differences at time points 1 and 3 are statistically significant). Additionally, when a glottal stop is produced by both genders, the vowel has lower pitch (there are of course no pitch measurements available for time point 3 as this time point occurs during the closure of the glottal stop). The notable differences on the basis of gender occur during the production of creaky voice or weak glottalization. In such productions, females must drastically lower their F_0 when glottalization occurs, whereas males do not.

Figure 2.16: Average pitch contours for GLOTTALIZED vowels by glottalization types and gender (production study 1, isolation)
modal voice = 'm'; weak glottalization = 'w'; creaky voice = 'c'; glottal stop = 'g'



This figure reinforces my claim that the semitone transform is suitable for pitch measurements taken during the production of modal voice but not for pitch measurements taken during the production of creaky voice. Furthermore, the fact that weak glottalization and creaky voice have a similar effect on pitch suggests that the use of the "weak glottalization" category is appropriate. Recall from §2.1.3.1 that this category refers to vowels that have an audible indication of creaky voice but whose only visual indication of glottalization is a brief lowering of intensity. Now we see that, in terms of pitch (as produced by females), vowels with weak glottalization are clearly distinguished from vowels with only modal voice.

Another tendency that is represented in the above graph is that, while high pitch in general is associated with GLOTTALIZED vowels, the dramatic decrease in pitch that is also associated with GLOTTALIZED vowels is not present when GLOTTALIZED vowel are produced with modal voice. This can be better understood if we consider the difference between minimum and maximum pitch values that are produced during vowel production; I refer to this value as the *pitch span* of the vowel. As shown in Table 2.12, the larger pitch spans occur in the weak glottalization and creaky voice categories regardless of which vowel shape is being produced.[26] This fact – that large pitch spans are directly correlated with glottalization type, and hence indirectly correlated with vowel shape – will have implications for the analysis of Chapter 5.

Table 2.12: Average pitch span (semitones) by gender, vowel shape, and glottalization type (production study 1, isolation)

		males	females
modal	GLOTTALIZED	3.8	2.9
	HIGH TONE	4.1	2.8
weak glottalization	GLOTTALIZED	5.8	6.0
	HIGH TONE	3.2	7.2
creaky voice	GLOTTALIZED	7.2	9.6
	HIGH TONE	9.0	11.3
glottal stop	GLOTTALIZED	4.2	6.8
	HIGH TONE	*n/a*	*n/a*

[26] The only deviation from this pattern is that HIGH TONE vowels produced with weak glottalization by males only have a somewhat small average pitch span of 3.2 semitones.

<u>Figure 2.17: Pitch contours of long vowels in Santa Elena by participant (production study 1)</u>
The following figure (continued on the next page) shows all pitch contours (top) and average pitch contours (bottom) as produced for GLOTTALIZED *(blue),* HIGH TONE *(red), and* LOW TONE *(green) by each participant from Santa Elena (gender and age of the speaker are given as the graph title). The pitch scale is Hertz.*

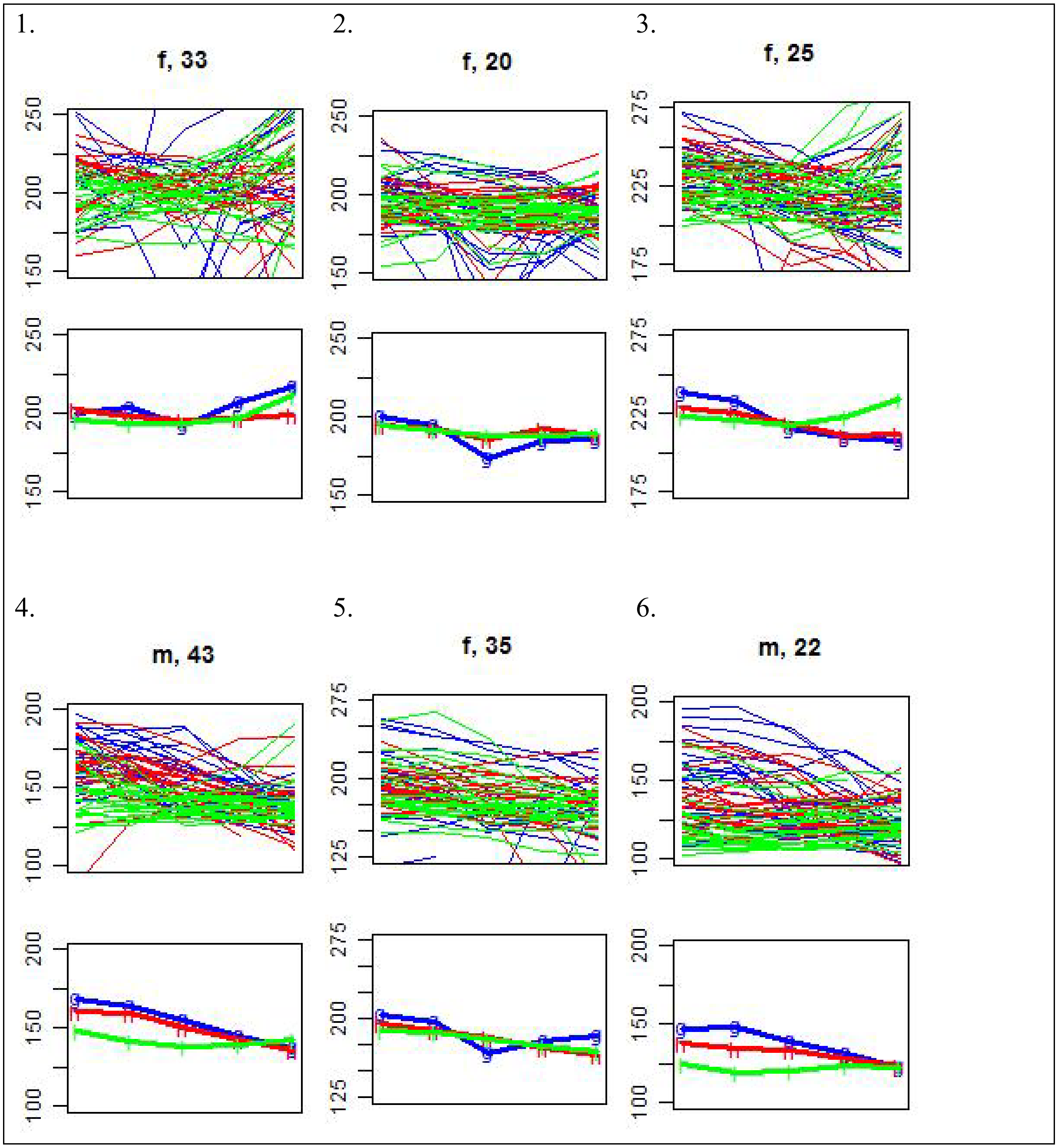


Figure 2.17 continued

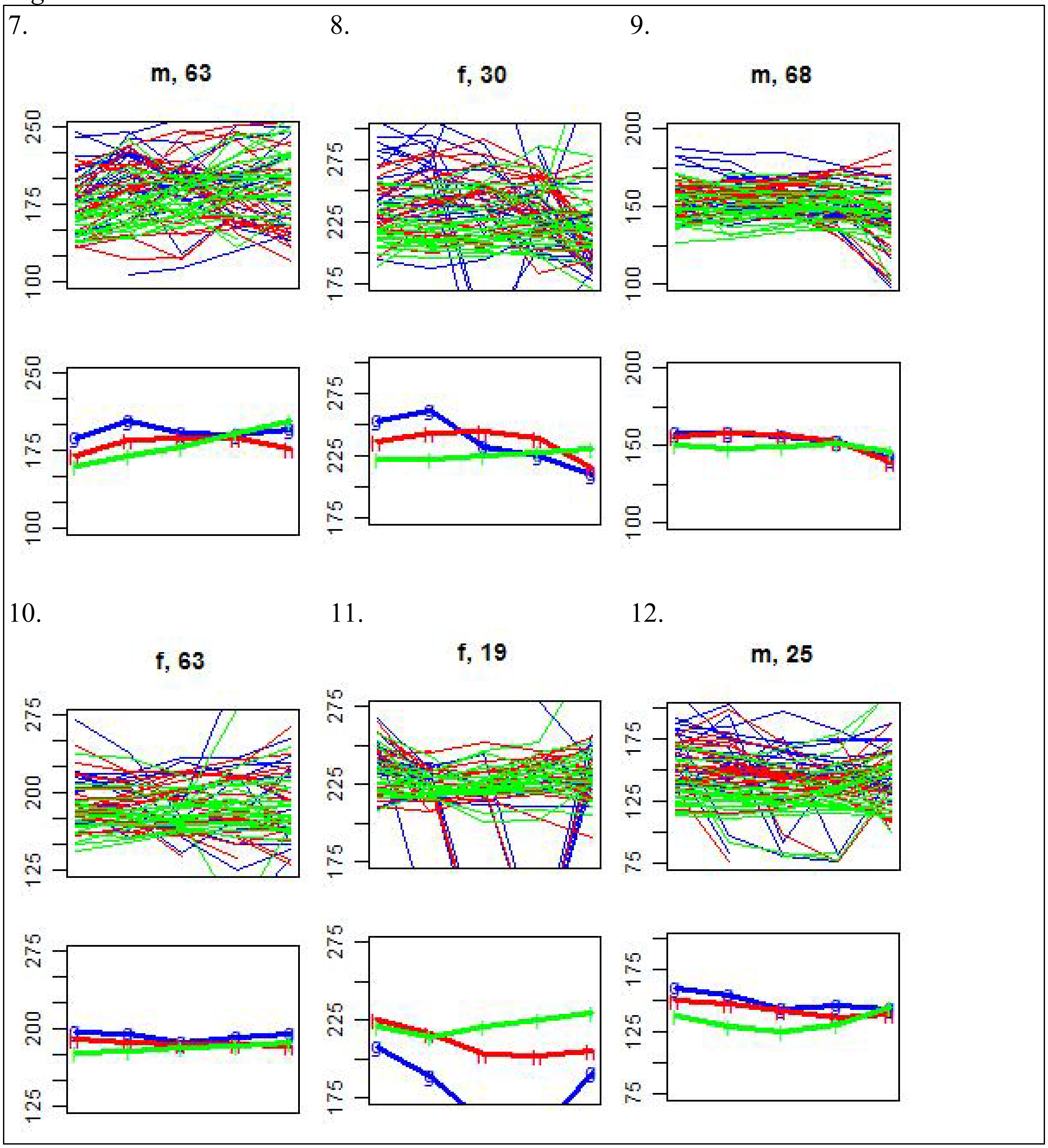


Fig. 2.17 shows all of the pitch contours for the long vowels produced in production study 1 by each participant in Santa Elena. This figure shows more inter-speaker variability in the production of pitch than was present with vowel length but less than was present with glottalization. Some speakers are very consistent in their productions of pitch (such as

participant 6), while others are less consistent. Additionally, not all speakers clearly distinguish GLOTTALIZED, HIGH TONE, and LOW TONE vowels by pitch contour (such as participants 9 and 10). Finally, not all speakers produce a falling contour (on average) for the GLOTTALIZED vowel (such as participants 1, 7, and 10).

There is no reason to assume that the idiolectal variation demonstrated by this data is particular to Yucatec Maya. It is clear, then, that in order to develop an accurate picture of the phonetic properties of any language, the larger number of speakers measured, the better. If, for example, only participants 1, 7, and 10 had been recorded, the production of pitch as described in this paper would be much different. When working with less-commonly studied languages, which are also often endangered, it is of course worthwhile to perform any phonetic analysis that we can, even with limited numbers of speakers, but it is also the case that every effort should be made to include multiple speakers so that idiolectal properties are not described as properties of the language as a whole.

That said, the linguist must generalize in order to make progress in our understanding of human language. Even though I have pointed to idiolectal phonetic variation, that variation occurs within a single speech community. Children are born into such speech communities, and they generalize over similar data in order to develop a working grammar. Idiolectal variation should not completely obstruct our task of describing a language's grammar (and not just a person's grammar), however, in order to minimize the possibility of idiolectal variation skewing our understanding of a language's grammar, the task of generalizing should include data from as many individual speakers from a given language community as possible. This is the best way to mimic the child language learner, who acquires a grammar on the basis of input from multiple sources.

We will now move on to the results of production study 2 with respect to pitch. In order to simplify the display, I will ignore the SHORT vowels and focus on the long vowels only, which are shown in Fig. 2.18.[27] The statistically significant differences among the pitch values of each vowel shape by time point and location are given in Table 2.13. The LOW TONE vowel is consistently produced with slightly falling contour, except in frame sentence C (*phrase-final (post-H))*, where it has a steeply falling contour. The GLOTTALIZED vowel always has a falling contour, though there is variation in how steeply it falls and where its peak occurs (either time point 1 or time point 2). The pitch contours of the HIGH TONE vowel, on the other hand, are quite variable. This vowel shape has a rising contour in context B (*phrase-medial*), a falling contour in context C and D (both *phrase-final*), and a peak in the middle in context A (*phrase-initial)*. As discussed in §1.2.4, Gussenhoven and Teeuw (2008) show that HIGH TONE vowels have a falling contour in phrase-final position but a rising contour in non-phrase-final position. The contours in Fig. 2.18 are in partial agreement with this, as HIGH TONE vowels do not have a clear rising or falling contour in context A (*phrase-initial*) and have a falling contour in context C (*phrase-final (post-H))*.

[27] The SHORT vowels do not show anything surprising in terms of pitch; their pitch contours are generally fairly level and produced with values in the middle of the speaker's range.

Figure 2.18: Average pitch contours for each frame sentence (production study 2)
GLOTTALIZED = 'g'; HIGH TONE = 'h'; LOW TONE = 'l'

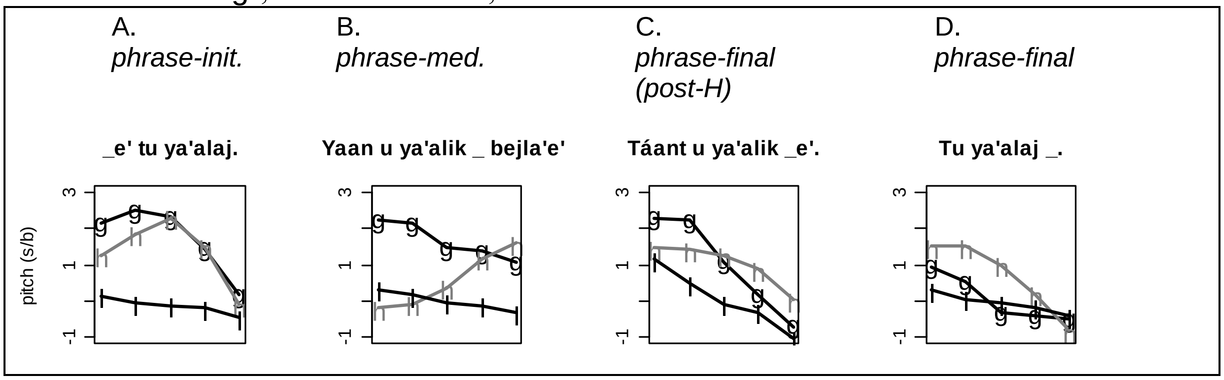


Table 2.13: Statistical analysis of pitch at each time point among vowel shapes by frame sentence (production study 2)

		statistically significant differences ($p < .05$, using a mixed linear regression model to account for multiple observations within subjects)
A	1	GLOTTALIZED > HIGH TONE > LOW TONE
	2	GLOTTALIZED > HIGH TONE > LOW TONE
	3	GLOTTALIZED, HIGH TONE > LOW TONE
	4	GLOTTALIZED, HIGH TONE > LOW TONE
	5	GLOTTALIZED > LOW TONE
B	1	GLOTTALIZED > HIGH TONE, LOW TONE
	2	GLOTTALIZED > HIGH TONE, LOW TONE
	3	GLOTTALIZED > HIGH TONE, LOW TONE
	4	GLOTTALIZED, HIGH TONE > LOW TONE
	5	HIGH TONE, GLOTTALIZED > LOW TONE
C	1	GLOTTALIZED > HIGH TONE, LOW TONE
	2	GLOTTALIZED > HIGH TONE > LOW TONE
	3	GLOTTALIZED, HIGH TONE > LOW TONE
	4	HIGH TONE > GLOTTALIZED, LOW TONE
	5	HIGH TONE > GLOTTALIZED, LOW TONE
D	1	HIGH TONE > LOW TONE
	2	HIGH TONE > LOW TONE, GLOTTALIZED
	3	HIGH TONE > LOW TONE, GLOTTALIZED
	4	*no significant differences*
	5	*no significant differences*

Gussenhoven and Teeuw (2008) claim that both the HIGH TONE and GLOTTALIZED vowels (which both bear a high tone marker) are downstepped after another HIGH TONE or GLOTTALIZED vowel. However, Gussenhoven and Teeuw only look at contours of HIGH TONE and GLOTTALIZED vowels averaged together; they do not present relevant data for each vowel shape separately. In my data there is no evidence that GLOTTALIZED vowels trigger or undergo downstep. For example, compare context A (*phrase-initial*) to context B (*phrase-medial*, where the target word follows a GLOTTALIZED vowel). GLOTTALIZED vowels have significantly higher pitch than LOW TONE vowels and the same falling contour in both contexts.

In context C (*phrase-final (post-H))*, HIGH TONE vowels start with the same pitch as LOW TONE vowels, though the contour of HIGH TONE vowels does not fall as rapidly as with LOW TONE vowels. In this context, the phrase begins with a HIGH TONE vowel, and hence the initial lower pitch of the target HIGH TONE vowels could represent some effect of downstep.[28] However, LOW TONE vowels are also irregular in context C in that they have a sharply falling contour instead of a slightly falling contour. It is possible that the initial HIGH TONE vowel of the sentence conditions both a lower initial pitch in the following HIGH TONE vowels and a higher initial pitch in the following LOW TONE vowels. More work is needed before we can determine the exact nature of the effect of HIGH TONE vowels on following HIGH TONE or LOW TONE vowels.

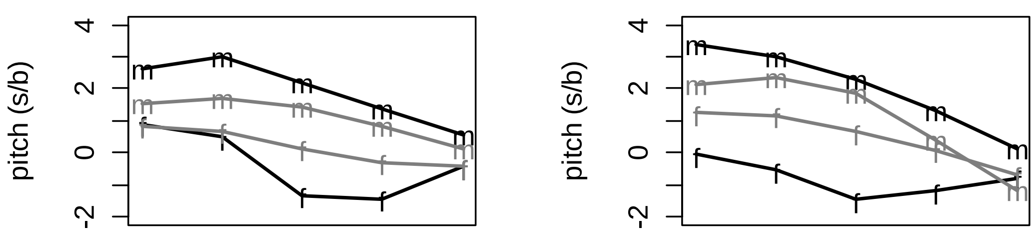
Pitch in context D (*phrase-final*) most closely resembles the results of production study 1, which is to be expected as isolated words are also phrase-final words. There is one notable difference, and that is that the initial pitch of HIGH TONE vowels is higher than the initial pitch of GLOTTALIZED vowels in context D. In order to better understand the differences and similarities between the pitch contours of production study 1 and the pitch contours of context D in production study 2, Fig. 2.19 shows a side-by-side comparison of

[28] While the peak of the HIGH TONE vowels is about at the same value in sentence C as in sentence B and D, we should be cautious about comparing absolute values across sentences, as it appears the effects of declination have not been completely factored out by the context-dependent semitone transform (see §2.1.3.2). Thus, if we compare the HIGH TONE contour to the LOW TONE contour within each context, we see these two contours are the closest in context C.

pitch by gender for both production studies. Here we see that males are fairly consistent in both contexts: GLOTTALIZED vowels start with higher pitch than HIGH TONE vowels and both vowel shapes have a falling contour. Females, on the other hand, produce HIGH TONE and GLOTTALIZED vowels with the same initial pitch in production study 1 but produce HIGH TONE vowels with higher initial pitch in production study 2. In production study 2, GLOTTALIZED vowels have relatively low pitch throughout, which of course drags that average down when all speakers are averaged together. Again, we are seeing the different effects of creaky voice on pitch as a consequence of gender. We will return to this result in Chapter 5 when it is time to develop a production grammar that accounts for this data.

Figure 2.19: Pitch contours for HIGH TONE and GLOTTALIZED vowels by gender (both production studies)
males = ‘m’; females = ‘f’; HIGH TONE = gray; GLOTTALIZED = black

a. isolation (production study 1) b. phrase-final (production study 2)



The results of production study 2 indicate that, even though the pitch contour of HIGH TONE vowels is variable, it is always distinguishable from the pitch contours of LOW TONE and GLOTTALIZED vowels. Additionally, these results show that target words in phrase-final position are produced in a manner similar to target words in isolation.

2.4.4 Local Summary

In Santa Elena, long vowels (GLOTTALIZED, HIGH TONE, LOW TONE) are almost twice as long as SHORT vowels in all contexts (produced in isolation or in frame sentences). There is a slight effect of phrase-final lengthening for the long vowels. Glottalization is produced with the GLOTTALIZED vowels 53% of the time when target words are spoken in isolation. When glottalization is produced, there is preference for the "creaky middle" type. When target words are produced in frame sentences, glottalization is more likely to be produced as the target word moves toward the end of the sentence.

All vowel shapes have a unique pitch contour in Santa Elena (as spoken in isolation in production study 1): the GLOTTALIZED vowel starts with high pitch and has a falling contour, the HIGH TONE vowel starts with high pitch (but not as high as the GLOTTALIZED vowel) and has a less steeply falling contour, the SHORT vowel has level mid-high pitch, and the LOW TONE vowel has low pitch with a slightly rising contour. When target words are produced in frame sentences, the pitch contour of the HIGH TONE vowel is either rising or falling depending on context, while the contours of the other vowel shapes remain fairly stable across contexts. In general, the production of vowel shape in words in isolation is similar to the production of vowel shape in words in phrase-final position.

There is an interaction between creaky voice and pitch in the GLOTTALIZED vowels. When GLOTTALIZED vowels are produced without glottalization, they have fairly steady high pitch. For males, glottalization does not strongly affect the pitch contour. For females, the production of creaky voice or weak glottalization causes a dramatic dip in F_0, resulting in a sharply falling pitch contour and hence a large pitch span. This is due to the fact that females and males produce creaky voice with the same F_0, a value that is in the normal range of pitch production for males but is much lower than the normal range of pitch production for females. This result will have implications for the theoretical analysis of Chapter 5.

In the next section we will examine the results for all other participants, some of whom behave similarly to the Santa Elena participants and some of whom do not. We will compare these speakers to the Santa Elena speakers in that section, and then the local summary in §2.5.4, and the chapter summary in §2.6 will outline the main findings with respect to dialectal differences.

2.5 Results for All Participants

In this section, I present the results of production studies 1 and 2 with respect to all participants. When applicable, I include the data from Santa Elena speakers in order to compare it to the data from other speakers, but the analysis in this section will focus on the other speakers. The data presented in this section will not be used for theoretical analysis in the later chapters, but is presented in order to document the wide array of dialectal variation in the production of vowel shape that has never before been discovered, or even speculated upon. Furthermore, because of the small numbers of participants from each town, especially in production study 2, the focus of this section is exploratory data analysis, though I will also show the results of hypothesis testing when appropriate.

2.5.1 Vowel Length

The results of production study 1 (isolation) indicate that, in all locations, long vowels are significantly longer than the SHORT vowel (Table 2.14). There is, however, some variation among the dialects as to which long vowel is the longest. As we saw in §2.4.1, in Santa Elena, HIGH TONE and GLOTTALIZED vowels are both longer than LOW TONE vowels, but in Mérida, the GLOTTALIZED vowel is significantly longer than either the LOW TONE or the HIGH TONE vowel, and in Sisbicchén it is the HIGH TONE vowel that is significantly longer than the other long vowels.

Table 2.14: Average vowel length (ms) by location (production study 1, isolation)
Statistically significant differences are indicated in the last row (*p < .05, using a mixed linear regression model to account for multiple observations within subjects*.)

	Santa Elena *12 speakers*	Mérida *7 speakers*	Sisbicchén *5 speakers*
SHORT	119	91	130
LOW TONE	197	193	207
HIGH TONE	214	197	226
GLOTTALIZED	220	208	214
	H, G>L>S	G>H,L>S	H>G,L>S

The results from production study 2 point to even more dialect variation than was evident from production study 1, as shown in Table 2.15. In Santa Elena, the average durations of the different types of long vowels are remarkably consistent; Mérida shows a similar pattern, though GLOTTALIZED vowels tend to be longer than HIGH TONE vowels which tend to be longer than LOW TONE vowels. In Sisbicchén, Xocén, and Yax Che, on the other hand, HIGH TONE vowels have extremely long durations while LOW TONE vowels are shorter than the other long vowels, especially in Xocén and Yax Che. Vowel length in these locations also varies more as function of frame sentence.

Table 2.15: Average vowel length (ms) by location (production study 2, frame sentences)

		frame sentence A	frame sentence B	frame sentence C	frame sentence D
		phrase-init.	*phrase-med.*	*phrase-final (post-H)*	*phrase-final*
		__e' tu ya'alaj.	Yaan u ya'alik __ bejla'e'.	Táant u ya'alik __e'.	Tu ya'alaj __.
Mérida *3 speakers*	SHORT	100	96	109	101
	LOW TONE	169	147	173	178
	HIGH TONE	164	158	168	187
	GLOTTALIZED	166	164	165	194
Santa Elena *14 speakers*	SHORT	113	117	112	108
	LOW TONE	179	169	176	187
	HIGH TONE	179	170	177	186
	GLOTTALIZED	172	176	176	186
Sisbic-chén *4 speakers*	SHORT	109	119	111	120
	LOW TONE	165	167	181	210
	HIGH TONE	175	195	189	233
	GLOTTALIZED	164	170	164	196
Xocén *3 speakers*	SHORT	107	99	100	86
	LOW TONE	174	146	172	143
	HIGH TONE	198	193	190	211
	GLOTTALIZED	193	167	172	186
Yax Che *2 speakers*	SHORT	127	132	126	135
	LOW TONE	164	162	173	164
	HIGH TONE	183	208	190	225
	GLOTTALIZED	163	176	166	185

The general picture that emerges from Table 2.15 is that there are two distinct patterns in the production of vowel length, with speakers from Mérida and Santa Elena demonstrating one pattern, while speakers from Sisbicchén, Xocén, and Yax Che demonstrate the other. Recall that Mérida and Santa Elena are located on the western side of the peninsula and that Sisbicchén, Xocén, and Yax Che are located quite close to each other on the eastern side of the peninsula (see map, Fig. 1.1). Thus, the production of vowel length presents the first type of evidence (and we will see further evidence in §2.5.3) for grouping eastern towns as one dialect and western towns as another. The differences between the two groups are explicitly presented in Fig. 2.20. Here we see that the long vowels in the western dialect are all roughly the same length regardless of sentence context, whereas in the eastern dialect, HIGH TONE vowels are always the longest and that phrase final lengthening occurs with all long vowels. The raw numbers used to create Fig. 2.20 are given in Table 2.16.

Figure 2.20: Average vowel length (ms) for two dialect groups (production study 2, frame sentences)
western towns (Mérida, Santa Elena) = black
eastern towns (Sisbicchén, Xocén, Yax Che) = gray

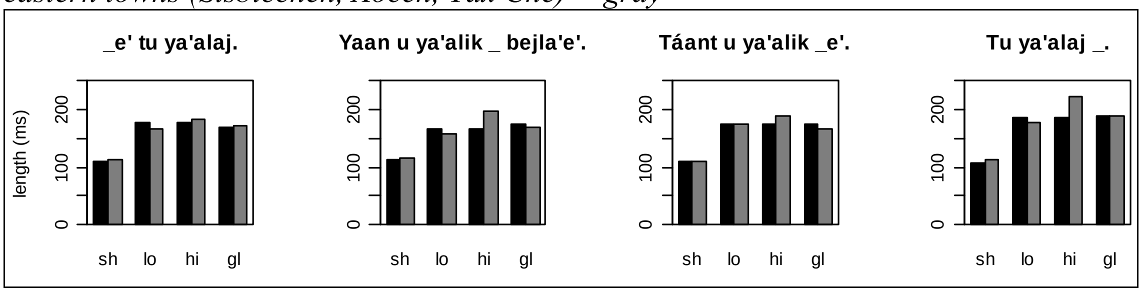


Table 2.16: Average vowel length (ms) for two dialect groups (production study 2, frame sentences)

		frame sentence A	frame sentence B	frame sentence C	frame sentence D
		phrase-init.	*phrase-med.*	*phrase-final (post-H)*	*phrase-final*
		__e' tu ya'alaj.	Yaan u ya'alik __ bejla'e'.	Táant u ya'alik __e'.	Tu ya'alaj __.
west *17 speakers*	SHORT	110	113	111	107
	LOW TONE	177	165	175	186
	HIGH TONE	177	168	175	186
	GLOTTALIZED	171	174	174	188
east *9 speakers*	SHORT	112	115	111	112
	LOW TONE	168	159	176	178
	HIGH TONE	184	197	190	224
	GLOTTALIZED	174	171	167	190

In summary, length is a contrastive feature for all speakers of Yucatec Maya, but for speakers in the east (Sisbicchén, Xocén, Yax Che), HIGH TONE vowels are always the longest. This result will be relevant for the discussion of pitch in §2.5.3.

2.5.2 Glottalization

In §2.4.2, I discussed the marginal production of glottalization in the non-GLOTTALIZED vowels, but in this section I will focus only on the GLOTTALIZED vowels. In production study 1 (isolation), the effect of location is significant, as shown in Table 2.17. Mérida speakers are more likely to produce glottalization and more likely to produce a full glottal stop than speakers from Santa Elena or Sisbicchén. Again though, as discussed in §2.4.2, there is a high degree of inter-speaker variation in the production of glottalization. For example, one speaker from Mérida produced a full glottal stop 36% percent of the time, while nine speakers (at least one from each location) never produced a glottal stop.

Table 2.17: Type of glottalization by location (GLOTTALIZED vowels only, production study 1)

	Santa Elena *n = 300*	Mérida *n = 175*	Sisbicchén *n = 125*
full glottal stop	6.7%	16.7%	2.4%
creaky	29.9%	61.4%	40.8%
weak glottalization	16.7%	7.4%	17.6%
modal voice	46.7%	14.5%	39.2%
Rao-Scott $\chi^2(6) = 22.9$, p = .0008; null hypothesis of equal distributions among locations			

In Table 2.18, we see the distribution of glottalization types by phrase position and location. We see a general decrease in the production of only modal voice and an increase in the production of creaky voice or a glottal stop as the target word moves toward the right edge of the sentence (as discussed in §2.4.2 and noted by Gussenhoven and Teeuw (2008)). Yax Che and Xocén show a smaller effect of sentence position.

Table 2.18: Distribution of glottalization types by sentence (GLOTTALIZED vowels only, production study 2, frame sentences)

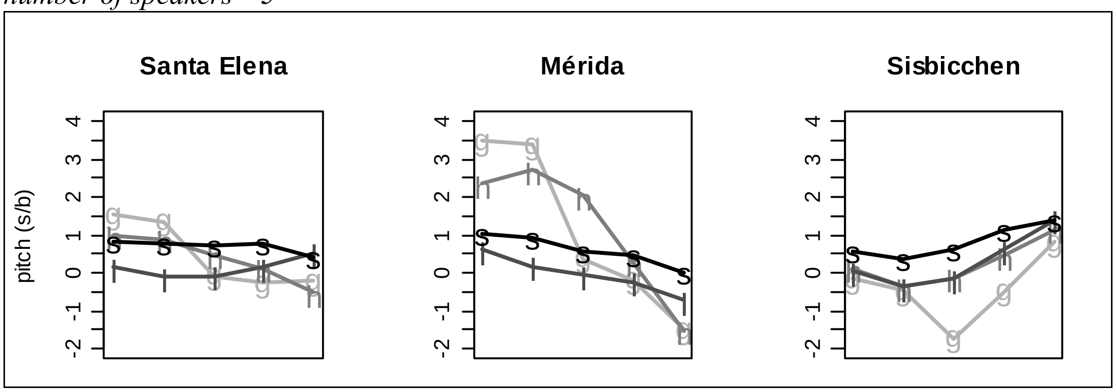
		frame sentence			
		A	B	C	D
		phrase-init.	*phrase-med.*	*phrase-final (post-H)*	*phrase-final*
		__e' tu ya'alaj.	Yaan u ya'alik __ bejla'e'.	Táant u ya'alik __e'.	Tu ya'alaj __.
Santa Elena n=504	glottal stop	0.0%	0.0%	0.8%	0.8%
	creaky	3.9%	13.5%	19.1%	25.5%
	weak glot.	4.0%	7.9%	10.3%	9.5%
	modal	92.1%	78.6%	69.8%	64.2%
Mérida n=108	glottal stop	0.0%	16.7%	3.7%	14.8%
	creaky	11.1%	37.0%	40.7%	44.4%
	weak glot.	25.9%	3.7%	11.1%	3.7%
	modal	63.0%	55.6%	44.4%	37.0%
Sisbic-chén n=144	glottal stop	0.0%	0.0%	0.0%	2.8%
	creaky	33.3%	47.2%	52.8%	72.2%
	weak glot.	22.2%	5.6%	8.3%	2.8%
	modal	44.4%	42.2%	38.9%	22.2%
Xocén n=108	glottal stop	11.1%	7.4%	18.5%	11.1%
	creaky	22.2%	44.4%	25.9%	37.0%
	weak glot.	11.1%	0.0%	3.7%	7.4%
	modal	55.6%	48.2%	51.9%	44.4%
Yax Che n=72	glottal stop	5.6%	5.6%	5.6%	5.6%
	creaky	33.3%	61.1%	50.0%	66.7%
	weak glot.	5.6%	11.1%	0.0%	11.1%
	modal	55.6%	22.2%	44.4%	16.7%

Even though Table 2.18 suggests that there may be dialectal patterns in the production of glottalization, the results of production study 2 show the same idiosyncrasies as we saw in production study 1. For example, only nine participants ever produced a full glottal stop, but at least one of these participants comes from each location. Additionally, three participants (two from Santa Elena and one from Mérida) never produced any form of glottalization. I conclude that glottalization is a feature of the GLOTTALIZED vowels in all dialects, and that the realization of glottalization may be loosely correlated with location, but that overall variation in the production of glottalization is idiolectal and not dialectal.

2.5.3 Pitch

The results with respect to pitch show some striking dialectal variation, such that it is possible a tonal merger has occurred for some speakers of Yucatec Maya. First, with respect to production study 1 (isolation), we see the average pitch contours for each vowel shape by location in Fig. 2.21. The pitch contours for Mérida and Santa Elena are very similar, but Sisbicchén presents a marked contrast to both of these locations. All vowel shapes in Sisbicchén show roughly the same rising contour (ignoring the super low pitch caused by creaky voice).

Figure 2.21: Average pitch contours by location (production study 1, isolation)
GLOTTALIZED = 'g'; HIGH TONE = 'h'; LOW TONE = 'l'; SHORT = 's'
for Santa Elena, number of speakers = 12; for Mérida, number of speakers = 7; for Sisbicchén, number of speakers =5



There are two noticeable differences between Santa Elena and Mérida. GLOTTALIZED and HIGH TONE vowels start with much higher pitch (in terms of actual value and in terms of value relative to LOW TONE vowels) in Mérida. While SHORT vowels have about the same pitch value in both locations, their pitch is more like the LOW TONE vowels in Mérida and more like the HIGH TONE vowels in Santa Elena. The statistically significant differences between the pitch of any two vowel shapes at each time point in Mérida is given in Table 2.19. The statistically significant differences for Mérida and Santa Elena (Table 2.11) are remarkably similar, and it seems the western towns produced pitch in the same manner.

Table 2.19: Statistical analysis of pitch among vowel shape in Mérida (production study 1, isolation)

	statistically significant differences ($p < .05$, using a mixed linear regression model to account for multiple observations within subjects)
1	GLOTTALIZED > HIGH TONE > SHORT, LOW TONE
2	GLOTTALIZED > HIGH TONE > SHORT > LOW TONE
3	HIGH TONE > SHORT, GLOTTALIZED, LOW TONE; SHORT > LOW TONE
4	SHORT, HIGH TONE > LOW TONE, GLOTTALIZED
5	SHORT > LOW TONE > HIGH TONE, GLOTTALIZED

Pitch contours in Sisbicchén certainly look different from those in Mérida or Santa Elena. As is made clear in Table 2.20, there are few significant differences between the pitch values for different vowel shapes for Sisbicchén. In general, SHORT vowels have a higher F_0 than the long vowels. Most interesting is the fact that there are no statistically significant differences in the pitch of "LOW TONE" and "HIGH TONE" vowels. We will evaluate the results from production study 2 before addressing this finding further.

Table 2.20: Statistical analysis of pitch among vowel shape in Sisbicchén (production study 1, isolation)

	statistically significant differences ($p < .05$, using a mixed linear regression model to account for multiple observations within subjects)
1	SHORT > GLOTTALIZED
2	SHORT > GLOTTALIZED, HIGH TONE, LOW TONE
3	SHORT > HIGH TONE, LOW TONE > GLOTTALIZED
4	SHORT > HIGH TONE, LOW TONE > GLOTTALIZED
5	SHORT > GLOTTALIZED

Again, we want to see how much inter-speaker variation there is in terms of pitch production, and so Figs. 2.22 and 2.23 show the pitch contours produced by each participant. The productions of speakers from both Mérida and Sisbicchén are quite consistent (within each location) with only a couple of exceptions, and so there does not seem to be as much idiolectal variation in these locations as there was in Santa Elena.

Figure 2.22: Pitch contours in Mérida by participant (production study 1, isolation)

The following figure (continued on the next page) shows all pitch contours (top) and average pitch contours (bottom) as produced for GLOTTALIZED *(blue),* HIGH TONE *(red), and* LOW TONE *(green) by each participant from Mérida (gender and age of the speaker are given as the graph title). The pitch scale is Hertz.*

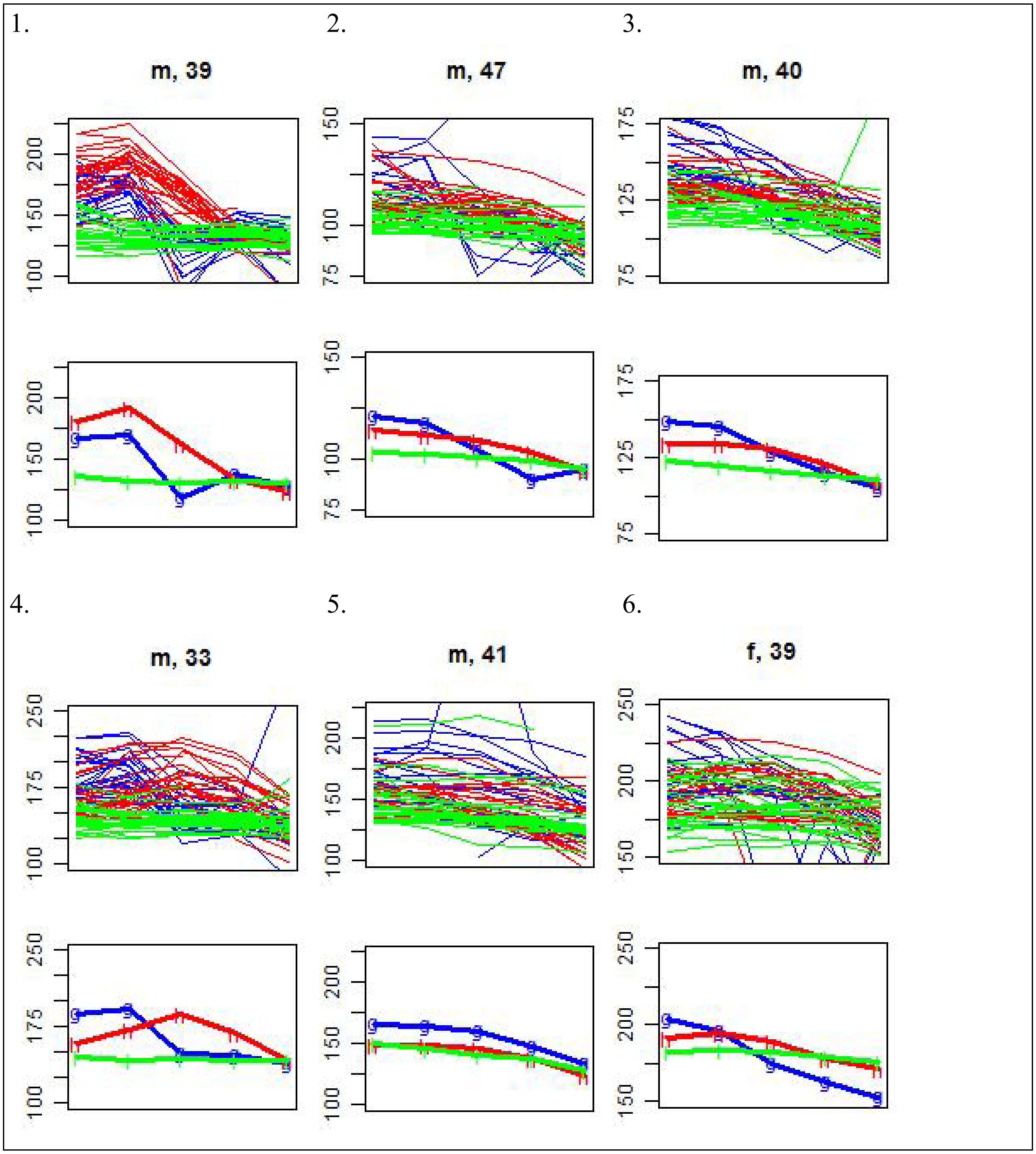


Figure 2.22 continued:

7.

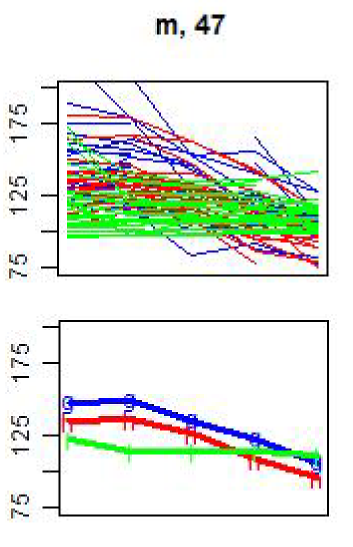
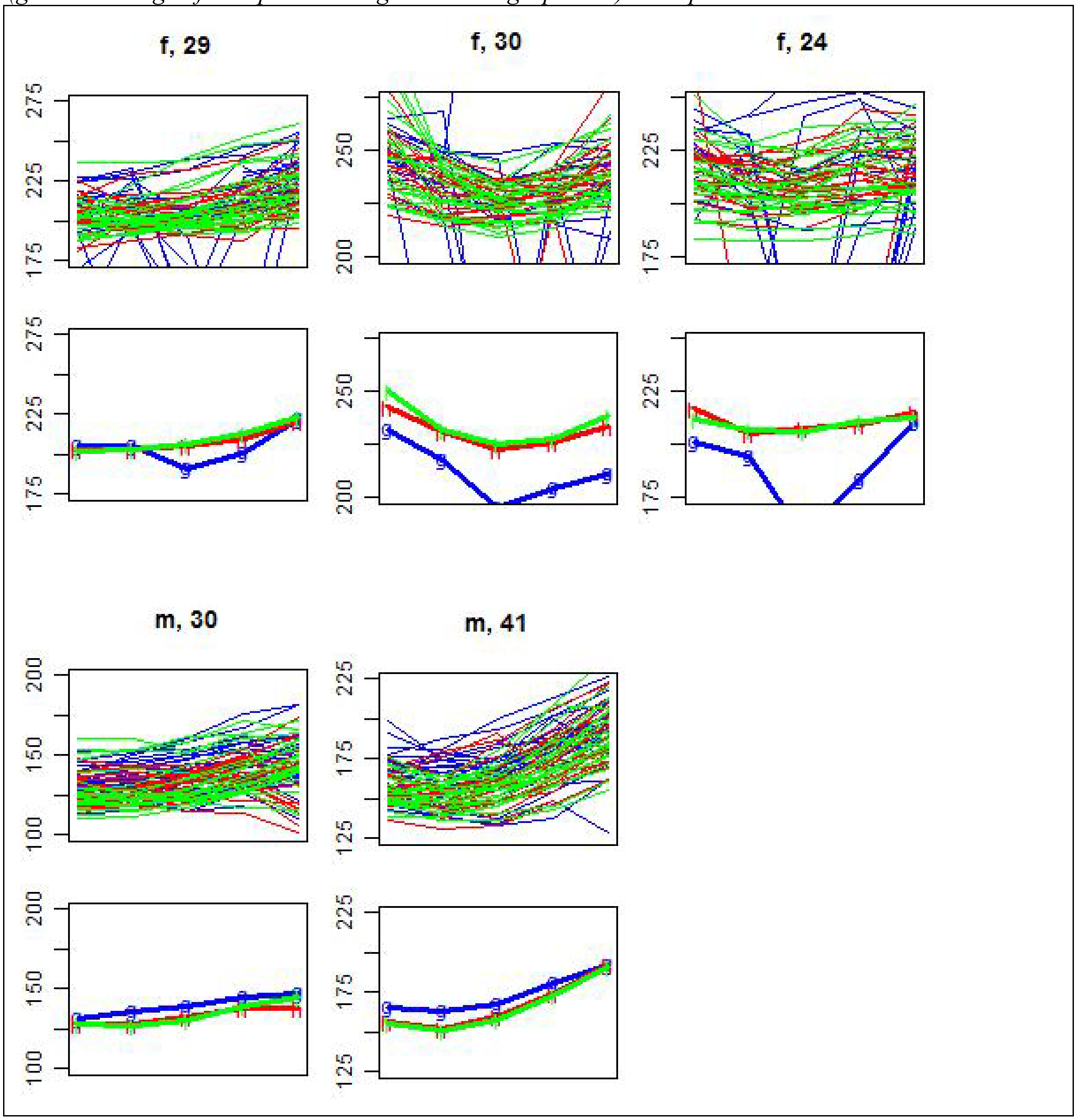
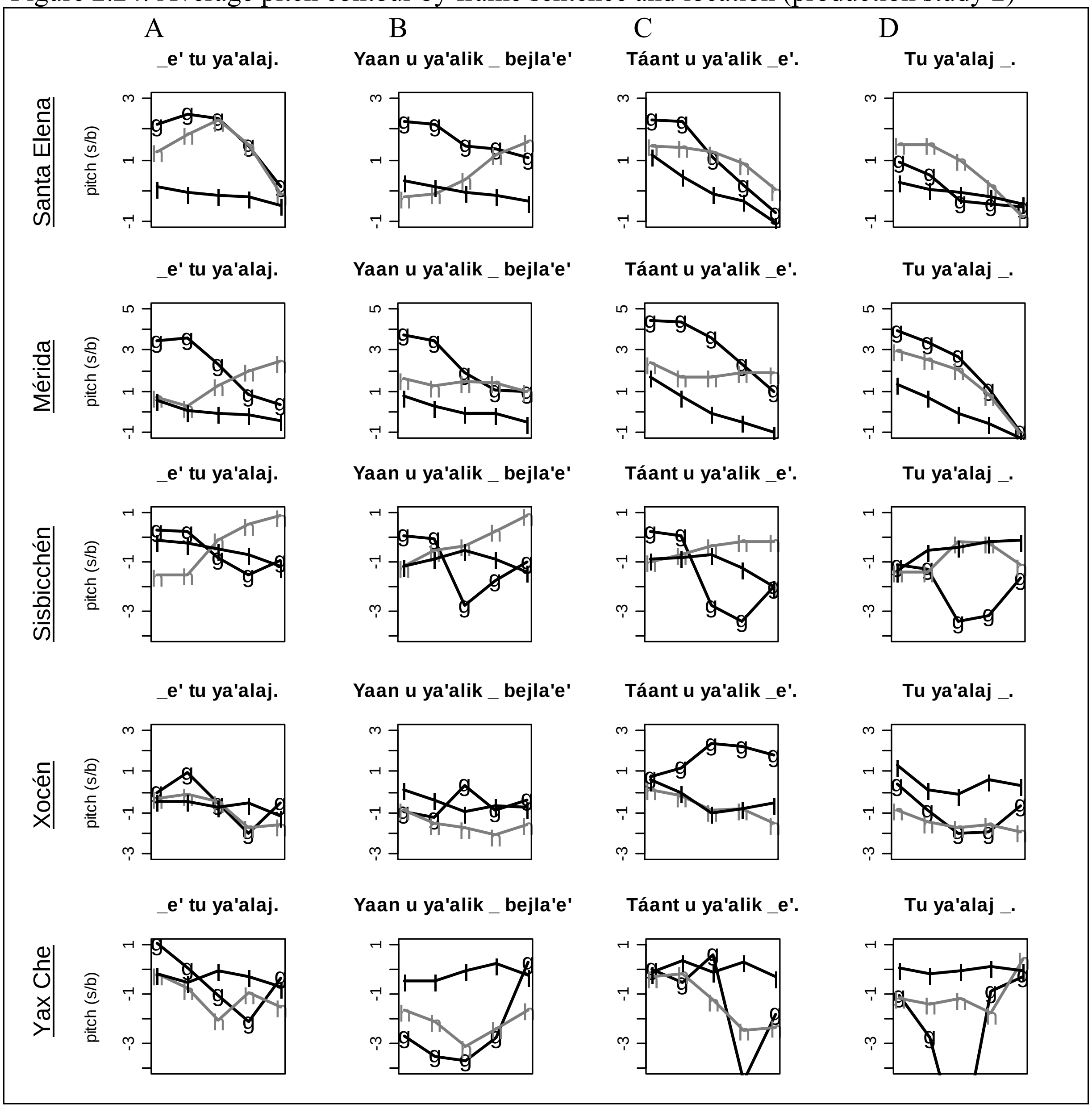


Figure 2.23: Pitch contours in Sisbicchén by participant (production study 1, isolation)
The following figure shows all pitch contours (top) and average pitch contours (bottom) as produced for GLOTTALIZED *(blue),* HIGH TONE *(red), and* LOW TONE *(green) by each participant from Sisbicchén (gender and age of the speaker are given as the graph title). The pitch scale is Hertz.*



In Fig. 2.24 we see the results of production study 2 with respect to pitch (in the long vowels only). We cannot conduct meaningful hypothesis testing with this data because of the small numbers of speakers. The tests we could perform would have to be based on the strong assumption that one speaker is representative of an entire population, and I have argued above that this assumption is unfounded. In the discussion that follows, I will first focus on the western cities (Santa Elena and Mérida) and then on the eastern cities (Sisbicchén, Xocén, Yax Che).

Figure 2.24: Average pitch contour by frame sentence and location (production study 2)



In Santa Elena and Mérida, LOW TONE vowels are produced with either steady or falling pitch, and HIGH TONE vowels tend to have a pitch peak that is higher than in LOW TONE vowels. Some possible exceptions are context C (*phrase-final (post-H)*, Santa Elena and

Mérida) and context B (*phrase-medial*, Mérida only). As discussed in §2.4.3, Gussenhoven and Teeuw (2008) claim that the initial HIGH TONE vowel in context C should trigger downstep, and so even though HIGH TONE vowels retain their higher pitch for longer than the LOW TONE vowels, it is possible that their initial pitch is lower (relative to the LOW TONE vowels) because of the preceding HIGH TONE vowel at the beginning of the sentence. The fact that HIGH TONE vowels do not have a rising contour in context B in Mérida is quite surprising. One possible explanation is the fact that vowels are shorter in this context (Table 2.14), and so there may not be enough time in the production of the vowel to reach a high pitch peak, but that would not explain why this is only the case with Mérida speakers and not those from Santa Elena.

In the other frame sentences, HIGH TONE vowels show some sort of contour in Mérida and Santa Elena, but the shape of that contour differs by context. There is clearly a falling contour for both locations in context D (*phrase-final*). Context A (*phrase-initial*) conditions a clear rising contour in Mérida, while context B (*phrase-medial*) conditions a clear rising contour in Santa Elena. The results for context A in Santa Elena and context B in Mérida do not fit neatly into the rising vs. falling dichotomy. Based on Gussenhoven and Teeuw's results we expect rising contours in these positions, as they are non-phrase-final. For sentence A in Santa Elena, HIGH TONE vowels do have a late rise, but it is unclear why pitch then falls at the end of the vowel. Despite these anomalies, the general pattern is in agreement with Gussenhoven and Teeuw: HIGH TONE vowels are produced with a falling contour in phrase-final position and with a rising contour in non-phrase-final position in the western cities.

The data for the cities in the east is quite different. Recall that production study 1, which only reported on participants from Sisbicchén (and not Xocén or Yax Che), showed no distinction between the pitch contours of HIGH TONE and LOW TONE vowels, and that each vowel shape was produced with a rising contour for speakers from Sisbicchén.

In Fig.2.24, we see that HIGH TONE and LOW TONE vowels do not always have the same contour in Sisbicchén: HIGH TONE vowels have rising contours in contexts A, B, and C, while LOW TONE vowels have falling contours in these sentences. In context D (*phrase-final*), however, HIGH TONE and LOW TONE vowels have roughly the same rising contour. This data suggests that, in Sisbicchén, the contrast between HIGH and LOW TONE vowels is neutralized in phrase final position (including monosyllabic utterances, as documented by production study 1), but is present – and realized as a distinction between falling and rising contours – in other positions. It is interesting that, referring back to Table 2.15, the biggest numeric difference between the lengths of LOW TONE and HIGH TONE vowels occurs in sentence D for speakers from Sisbicchén – exactly the context where the pitch differences are neutralized. Sisbicchén also shows the same falling contour for GLOTTALIZED vowels regardless of context, where time points 3 and 4 clearly show the effect of creaky voice on pitch.

In Xocén, we see almost no difference between HIGH TONE and LOW TONE vowels as well as among sentence contexts. The notable exception is frame sentence D (*phrase-final*), where LOW TONE vowels are produced with higher pitch than HIGH TONE vowels. In Yax Che, LOW TONE vowels tend to be produced with higher pitch than HIGH TONE vowels, but the pitch contours are inconsistent among productions. Furthermore, in sentence A (*phrase-initial* and sentence D (*phrase-final*), the pitch contours of HIGH TONE and LOW TONE vowels are quite similar.

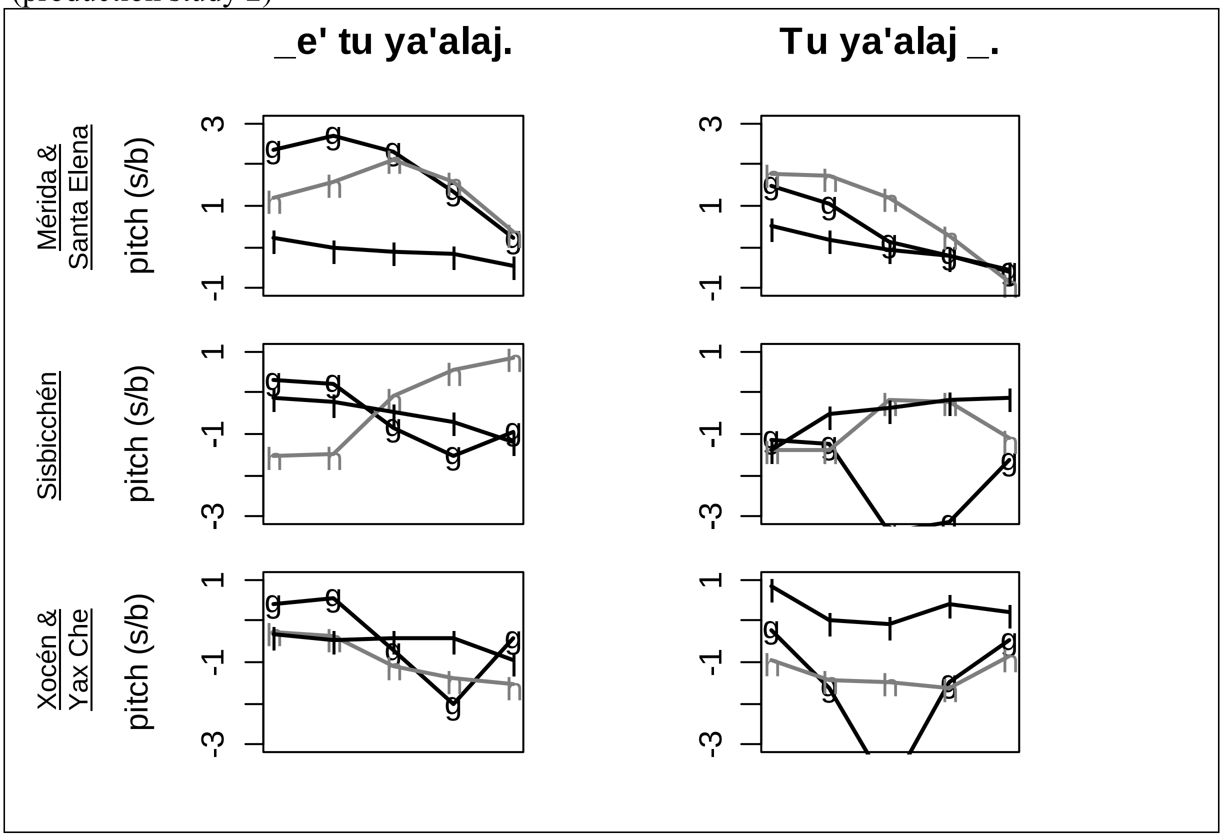
Though it was originally the results from Sisbicchén that suggested that tone is a dialectal feature of Yucatec Maya, it now appears that tone at least still marginally present in Sisbicchén, but not present at all in the two other eastern cities of Xocén and Yax Che.

In Xocén, we see a slightly falling contour for GLOTTALIZED vowels in contexts A, B, and D, but a rising contour in sentence C (*phrase-final (post-H))*. In Yax Che, the production of pitch in GLOTTALIZED vowels is considerably different among the sentences, with sharp drops in pitch due to creaky voice in sentences C and D (both *phrase-final*), a falling contour in sentence A (*phrase-initial*) and a rising contour in sentence B (*phrase-medial*). One factor to keep in mind is that, as demonstrated in §2.2, creaky voice has more of an impact on the production of pitch in females, and that all participants from Yax Che and all but one from Sisbicchén are female. This could explain the dramatic dips in the pitch contours in with these locations. If we abstract away from this, we see that Sisbicchén produces pitch contours with GLOTTALIZED vowels that are very similar to those of the western cities, but that Yax Che and Xocén do not. Again, it seems that Sisbicchén does not deviate from the standard view of pitch in Yucatec Maya as much as production study 1 suggested, but that Xocén and Yax Che produce pitch contours that are not expected.

At this point we can make the following generalizations. Speakers from Mérida and Santa Elena behave similarly by producing consistent falling pitch contours for GLOTTALIZED vowels, falling pitch contours for HIGH TONE vowels in phrase-final position, rising contours for HIGH TONE vowels in non-phrase-final position, and (mostly) level pitch for LOW TONE vowels. Speakers from Sisbicchén produce falling pitch contours for GLOTTALIZED vowels, falling pitch contours for LOW TONE vowels (except in sentence D) and rising contours for HIGH TONE vowels. Speakers from Xocén and Yax Che do not clearly distinguish pitch in

HIGH TONE and LOW TONE vowels and produce a variety of contours in GLOTTALIZED vowels. These generalizations point to three distinct strategies for the production of pitch on the basis of location and to a distinction between pitch contours in phrase-final and non-phrase final position. Fig. 2.25 summarizes these generalizations by looking at pitch as produced by three different dialect groups and in two different frame sentences.

Figure 2.25: Average pitch contours by three location groups and two frame sentences (production study 2)



2.5.4 Local Summary

The data in §2.5 has indicated a variety of dialectal differences in the production of vowel shape in Yucatec Maya. With regard to vowel length, all locations distinguish long vowels from SHORT vowels in all contexts, but the longest long vowel varies by location (with GLOTTALIZED vowels being the longest in the western cities and HIGH TONE vowels being the longest in the eastern cities). Glottalization is a feature of the GLOTTALIZED vowels in all locations. Glottalization is most often realized as creaky voice, though there is idiolectal and not necessarily dialectal variation in its production.

Both sentence context and location influence the production of pitch. Speakers from Mérida behave similarly to those from Santa Elena, and thus it seems that in all respects speakers from these two towns behave as one dialect group. In Sisbicchén, HIGH TONE and LOW TONE vowels are not distinguished by pitch in phrase-final position but are distinguished by a rising vs. falling contour in other positions. Speakers from Xocén and Yax Che provide no consistent evidence for a tonal contrast. What is most interesting about this result is that HIGH TONE vowels are significantly longer than LOW TONE vowels in these eastern cities where pitch is either not phonemic or where the contrast is neutralized in certain contexts. This data suggest that, while a tonal merger may have occurred in some locations, a phonemic merger has not. It would be valuable for future work to investigate this phenomenon further, since I was not able to record enough speakers in the east to provide conclusive data on the topic.

2.6 Chapter Summary

This chapter has explained the methodologies used for both production experiments and presented the results of these studies with respect to vowel length, pitch, and glottalization. In the discussion of §§2.4 and 2.5, I have interpreted the data as I found appropriate, but I have also tried to supply enough raw data so that others may make their own interpretations.

The evidence presented in the preceding sections has suggested that there are currently at least two different dialects of Yucatec Maya with regard to the production of vowel shape. In the two cities on the western side of the peninsula – Mérida and Santa Elena

– the contrasting properties of vowel shape are those that are described in the literature (see §1.2.4). Long vowels are longer than short vowels, both GLOTTALIZED vowels and HIGH TONE vowels start with initial high pitch in phrase-final position, whereas in non-phrase-final position, the GLOTTALIZED vowels still have a falling contour but the HIGH TONE vowels have a rising contour. LOW TONE vowels are produced with steady or slightly falling low pitch (though they also have a slightly rising contour in some contexts for Santa Elena speakers). The only result with regard to the western cities that directly contradicts the claims of previous literature is that glottalization is most commonly realized as creaky voice, and only very rarely as a full glottal stop.

The eastern cities of Sisbicchén, Xocén, and Yax Che produce vowel shape in a way that is different from the western cities. Glottalization is still canonically produced as creaky voice, but vowel length and pitch are quite different. HIGH TONE vowels are consistently longer than the other long vowels, but pitch is not consistently distinguished between HIGH TONE and LOW TONE vowels. This data suggests the possibility that, while a tonal merger has occurred, a three-way length distinction has emerged in its place. More work is needed in this area due to the low numbers of speakers from the east that participated in the production studies.

Based on the data presented above, I propose the phonological representations given in Table 2.21 for the vowel shapes in the western dialect (Mérida and Santa Elena) of Yucatec Maya (which were first provided in Table 1.2). I leave short vowels unmarked for tone at this level. I also assume that tonal markers occur with the first mora of a long vowel, and the second mora is unmarked with respect to tone. In phrase-final position, all long vowel shapes end with low pitch, indicating that either the phonological grammar inserts a low tone marker on the second mora or that the phonetic grammar conditions low pitch at the end of a phrase. By leaving the second mora unmarked for tone in the HIGH TONE vowels, the high tone marker is free to move into this position when in non-phrase-final position. I assume that the last portion of the GLOTTALIZED vowel is marked for creaky voice at the phonological level and that the phonetic grammar accounts for the various strategies of producing glottalization. Finally, I propose that the creaky voice s marked on the second mora of the GLOTTALIZED vowel and that this accounts for why high tone is never realized on the last portion of this vowel.

Table 2.21: Phonological representation of vowel shape in the western dialect

	Mérida/Santa Elena "western dialect"
SHORT	/v/
LOW TONE	/v̀v/
HIGH TONE	/v́v/
GLOTTALIZED	/v́v̰/

In the theoretical component of this dissertation (especially Chapter 5), I will focus on speakers from Santa Elena. Though it is clear that speakers from both Santa Elena and Mérida form a dialect group, there are still enough minor differences between the two locations that I find it preferable to use speakers from only one location.

3

PHONETIC DESCRIPTION OF YUCATEC MAYA: PERCEPTION OF PITCH AND GLOTTALIZATION

The production studies presented in Chapter 2 were designed to evaluate which acoustic cues differ among the vowel shapes of Yucatec Maya. We saw in that chapter that, in all dialects, the SHORT vowel is distinguished from the other vowel shapes by length, and the GLOTTALIZED vowel is distinguished from the other vowel shapes by the production of glottalization.

The use of pitch as a cue depends on dialect. In the western dialect (Mérida and Santa Elena), LOW TONE vowels have much lower pitch than the other long vowels. The GLOTTALIZED and HIGH TONE vowels, on the other hand, are produced with phonetic values for pitch that have a high degree of overlap. The GLOTTALIZED vowels tend to be produced with glottalization, but this only occurs half the time. When they are produced with modal voice, glottalization cannot be used to distinguish them from HIGH TONE vowels. Both of these vowel shapes have a falling pitch contour and are marked for high tone during the initial portion of the vowel. We did see evidence (see especially Fig. 2.21) that GLOTTALIZED vowels tend to be produced with higher initial pitch and a larger pitch span than HIGH TONE vowels. It seems then that, for the western dialect, no one cue may be perfect at distinguishing GLOTTALIZED from HIGH TONE vowels but that multiple cues might be integrated into this contrast.

For speakers from towns in the east (Sisbicchén, Xocén, and Yax Che) glottalization is the only cue that consistently differs between the HIGH TONE and GLOTTALIZED vowels. The pitch contours of these two vowel shapes vary by context, but the data from the production studies showed enough variation that it is hard to discern any consistent patterns.

In this chapter we focus on the perception of the HIGH TONE and GLOTTALIZED vowels of Yucatec Maya. Our goals are to determine 1) how successful listeners are at discriminating between these two phonemic categories, 2) what cues listeners use to discriminate, and 3) if the dialect variation that is present in production is mirrored in perception. Given that symmetries often occur between production and perception (as discussed in Chapter 1), I hypothesize that listeners will use all the cues that are differentiated in production as cues to perceiving the contrast between these two vowel shapes. This means that listeners from the east will use different strategies than listeners from the west.

In task 1, where the stimuli were tokens of natural speech, this hypothesis is confirmed. Listeners from the west use both pitch and glottalization in determining which word they heard, while listeners from the east are more likely to use glottalization alone. The results of task 2 are different; listeners from the west use glottalization alone, while listeners from the east use glottalization and, to some degree, pitch. The differences between the results of these two tasks could be related to the difference in stimulus quality (natural vs. manipulated tokens). I propose that listeners made a decision to change perceptual strategies in the face of unnatural stimuli and thus focused on the cue that is most likely to lead to successful comprehension.

This chapter is organized as follows. I first present the general methodology of the perception study in §3.1, and describe how the tasks performed by the participants are related to the Bidirectional Model in §3.2. Each task is then discussed separately. The specific methodology and the results for task 1 are provided in §3.3 and task 2 follows in §3.4. Within each task, the results for Santa Elena participants are presented first, followed by the results for all participants. In §3.5 I compare the results of the two tasks and argue that they can be accounted for with different perceptual strategies that are triggered by stimulus quality. Discussion and a chapter summary are provided in §3.6.

3.1 Methods Common to Both Tasks

Twenty-three participants, from the towns of Santa Elena, Mérida, Sisbicchén, Xocén, and Yax Che (see map in Fig. 1.1), took part in a perception study that involved two different tasks. All participants are fluent in Spanish in addition to Yucatec Maya; all speak Yucatec Maya in the home and in daily life (see Table 3.1 for more information). The perception study took place at the same time as production study 2 (Chapter 2), and all but one of the participants in the perception study also participated in production study 2. Because our focus in the analysis of Chapter 5 is the perception grammar of the Santa Elena dialect, the results for this dialect group are presented separately in §§3.3.2 and 3.4.2. The results for all speakers are addressed in §§3.3.3 and 3.4.3.

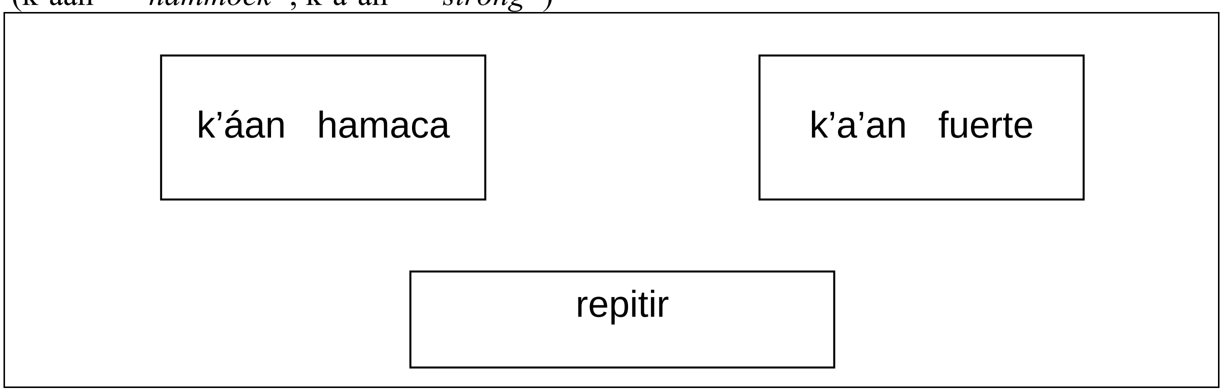
Table 3.1: Participant data

location	gender	ages	notes
Santa Elena	9 females	21, 21, 21, 23, 26, 31, 34, 38, 65	one fluent in English
	5 males	23, 43, 44, 64, 69	one fluent in English
Mérida	2 males	20, 25	originally from smaller town on the western side of peninsula
Sisbicchén	3 females	29, 31, 36	have lived in Valladolid
Xocén	2 males	20, 35	
Yax Che	2 females	21, 28	

Participants performed two different tasks. Each was a forced choice task such that 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 and Weenink 2006). The participant listened to each stimulus through headphones and then used a mouse to select the appropriate word on the computer screen (the stimuli and word choices are unique to each task and are discussed in §§3.3.1 and 3.4.1). For each choice on the computer screen, a word was displayed in both Yucatec Maya and Spanish, as shown below. 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.

Figure 3.1: Sample screen shown during perception experiments
(k'áan = *"hammock"*; k'a'an = *"strong"*)

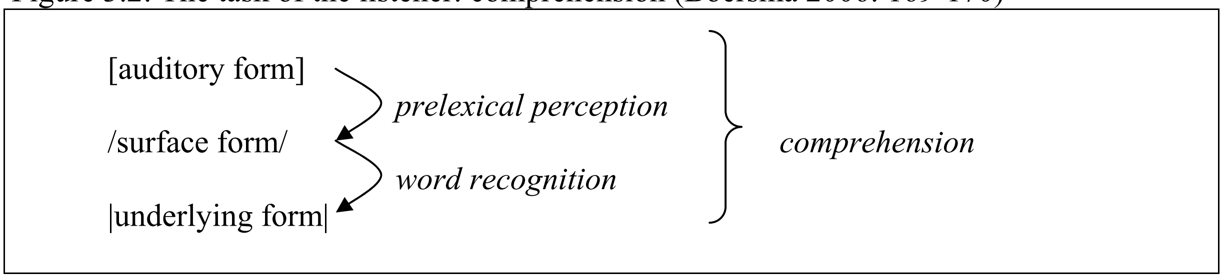


Before the actual tasks of the experiment began, participants went through a practice perception task. This was done to ensure both that the participant understood the instructions and that they were comfortable using a mouse to interact with the computer screen. If necessary, the participant was shown 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. Further details about the stimuli used for each task and the procedure of the task itself are provided as each task is discussed below.

3.2 The Perception Tasks in the Context of Bidirectional StOT

The components of the Bidirectional Model were briefly introduced in §1.1.3 (and they will be discussed in detail in the next chapter). Fig. 3.2 shows how comprehension is handled with this model.

Figure 3.2: The task of the listener: comprehension (Boersma 2006: 169-170)



Though the comprehension of speech requires the listener to complete both the processes of prelexical perception and of word recognition, Boersma (2006) discusses how the linguistic task of phoneme categorization may be accurately modeled as just the relation between an auditory form and a surface form, i.e. as prelexical perception. If the experimental design is "set up in such a way that the influence of the lexicon is minimized (p. 171)", then the auditory to surface form mapping alone can account for the process.

Both tasks of this perception study were designed to minimize the influence of the lexicon. The listener hears a word and must choose between two words that differ only by vowel shape, and the listener knows beforehand that this is the choice they will have to make. This means that listener can be expected to listen for cues as to which phoneme they heard (a GLOTTALIZED vowel or a HIGH TONE vowel), i.e. they are expected to categorize a particular phoneme that was produced. Furthermore, each task is monotonous, which is a factor that has been shown to result in the disappearance of lexicality effects on phoneme-monitoring responses (Cutler et al. 1987).

Thus, it is best to interpret the task of the participants in this study as one of prelexical perception, in which they hear an auditory form and map it onto a surface form. In the theoretical analysis of Chapter 5, I will account for how the auditory forms that each participant hears are mapped onto the phonological categories that they choose.

3.3 Task 1: Discrimination between Natural Tokens of HIGH TONE and GLOTTALIZED Vowels without Additional Context

The goal of this task is twofold: to determine if native listeners can differentiate between natural tokens of HIGH TONE and GLOTTALIZED vowels without any other context and to begin an investigation of which phonetic cues influence the choice of the listener. The first goal is not trivial given the degree of similarity in the production of both pitch and glottalization of these two vowel shapes. In both dialects, many GLOTTALIZED vowels are produced with modal voice throughout the duration of the vowel, which makes them indistinguishable from HIGH TONE vowels on the basis of glottalization, and the distributions of pitch values show a large amount of overlap between the two vowel shapes. Furthermore, in the western dialect, when GLOTTALIZED vowels are produced with modal voice, their pitch contours even more closely resemble the pitch contours of HIGH TONE vowels; it is usually when GLOTTALIZED vowels are produced with glottalization that they have really high initial pitch and a really large pitch span.

Because acoustic productions of these vowels resemble each other so closely, it is possible that a "near merger" has occurred. For example, Yu (2007) documents a case where two rising tones in Hong Kong Cantonese are produced with significantly different F_0 contours, but listeners cannot do better than chance at identifying which tone they heard. The question is then whether or not Yucatec Maya speakers can correctly distinguish HIGH TONE from GLOTTALIZED vowels when given no contextual clues.

3.3.1 Stimuli and Procedure

In production study 1 (Chapter 2), participants read target words in isolation, and the word list included one minimal pair where a HIGH TONE vowel contrasts with a GLOTTALIZED vowel: *k'a'an* [k'áa̰n] 'strong' vs. *k'áan* [k'áan] 'hammock'. This is a robust minimal pair in that both words are in everyday usage and there is no debate about the vowel shape of each lexical item. Furthermore, because *k'a'an* is an adjective, there are no morphosyntactic processes that change its vowel shape. As discussed in §1.2.8, *k'áan* does undergo a change in tone when it is possessed (*in k'aan* [k'àan] 'my hammock' (Lehmann 1998)), but this happens only in cases of possession and never when the word is used in isolation or unpossessed. Each unmanipulated production of *k'a'an* and *k'áan* as obtained in production study 1 was used as a stimulus for this task (48 stimuli total, 24 productions of each word). Participants heard each stimulus once and were forced to choose between *k'a'an* and *k'áan* as the word they thought they heard, as detailed above.

Recall that speakers from three different towns (Santa Elena, Mérida, and Sisbicchén) participated in production study 1 and that the data from speakers from Sisbicchén looks quite different from the other towns. With regard to the HIGH TONE and GLOTTALIZED vowels, speakers from Sisbicchén produced both with rising contours, which are quite unlike the falling contours of Santa Elena and Mérida. In this perception task, listeners heard all 48 stimuli, which come from speakers from all three towns. This means that listeners from the west heard tokens as spoken by speakers in the east and vice versa. I did not find any evidence in my data that listeners adjusted their perception strategy to accommodate the dialect of the speaker (see Table 3.6 below where the Santa Elena participants' responses are broken down by the dialect of the speaker of the stimulus). This could be because listeners were not given enough information to discern dialect, because (some of) the listeners were not familiar with different dialects,[29] or because, once this data set is divided into multiple dialect groups, there are not enough tokens to obtain significant results. In the following sections, I will not include the dialect of the speaker as a variable as there is currently no evidence that this influences the listener. There would certainly be reasons, though, for future research to investigate the effect of speaker dialect on perceptual strategy.

3.3.2 Results for Santa Elena

Participants were able to identify which word they heard on the basis of vowel shape at a rate better than chance, but just barely, as shown in Table 3.2.

Table 3.2: Task 1 responses by stimulus (Santa Elena only)
(The number in parentheses indicates the number of tokens.)

	stimulus	
	k'a'an (G)	*k'áan* (H)
resp. = GLOT	63% (212)	37% (124)
resp. = HIGH	37% (124)	63% (212)
Rao-Scott $\chi^2(1) = 16.1$, $p < .0001$		

If we look at the phonetic values of the different dimensions that were distinguished in production, we see that initial pitch, pitch span, and glottalization type all influence the choice of the participant in task 1 of the perception experiment. There is a significant effect of initial pitch (z = 3.4, p = .0008) and of pitch span (z = 4.6, p < .0001),[30] such that as the stimulus presents either a higher initial pitch or a larger pitch span, the listener is more likely to select the word with a GLOTTALIZED vowel. We can investigate more closely the effect of initial pitch and pitch span on the listeners' responses by categorizing the pitch scales.

[29] Each participant filled out a questionnaire which asked where they had lived, but it did not ask where they had traveled. From what I know of some of the participants, some have traveled widely and would certainly have spoken with native speakers from multiple parts of the peninsula and some have not traveled far and possibly would never have had personal experience with another dialect.

[30] Both tests 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 (initial pitch or pitch span) and the logit of the dependent variable (response).

Tables 3.3 and 3.4 show the percentage of times participants selected a glottalized vowel for six categories of initial pitch values and pitch spans. The categories used in these tables are defined in §5.2. The initial pitch category of -2 includes all pitch values that are less than or equal to -2; 0 includes all values less than or equal to 0 (and greater than -2), etc. The highest category (initial pitch = 8, pitch span = 12) includes all the values higher than the previous category.

Table 3.3: Task 1 responses by initial pitch (Santa Elena only)

	initial pitch category					
	-2	0	2	4	6	8
resp. = GLOT	52%	44%	38%	59%	79%	50%

Table 3.4: Task 1 responses by pitch span (Santa Elena only)

	pitch span category					
	2	4	6	8	10	12
resp. = GLOT	33%	41%	59%	63%	71%	84%

As shown in Table 3.5, participants are more likely to select a word with a GLOTTALIZED vowel if the stimulus contains any form of glottalization. A full glottal stop almost guarantees the selection of a GLOTTALIZED vowel, and, surprisingly, weak glottalization is correlated with more GLOTTALIZED vowel selections than creaky voice.

Table 3.5: Task 1 responses by glottalization type (Santa Elena only)

	glottalization type			
	ng	wg	cr	gs
resp. = GLOT	39%	68%	55%	93%

overall effect of glottalization: Wald $\chi^2(3)$ =21.6, p < .0001
mod vs. wg: Wald $\chi^2(1)$ = 20.7, p = <.0001
mod vs. cr: Wald $\chi^2(1)$ = 4.5, p = .034
mod vs. gs: Wald $\chi^2(1)$ = 13.4, p = .0003
wg vs. cr: Wald $\chi^2(1)$ = 3.9, p = .048
wg vs. gs: Wald $\chi^2(1)$ = 7.2, p = .007
cr vs. gs: Wald $\chi^2(1)$ = 10.9, p = .001
(standard errors are adjusted for multiple observations within subjects)

Because the production data on pitch showed gender differences when creaky voice was involved (when pitch is measured with the semitone transform), another question worth pursuing is how (or if) these differences are encoded in the grammar. In other words, are the production differences based on gender mirrored in perception? One possibility is that females have different production grammars and hence different perception grammars (according to the Bidirectional Model, see §4.3). This would mean that female listeners use different perceptual strategies. Another possibility is that all language users have the same grammar but that that grammar encodes information on gender of the speaker. This would mean that all listeners use different perceptual strategies depending on the gender of the speaker.

With regard to the first option, that listeners of different genders use different perceptual strategies, the results of task 1 provide no evidence of this. Table 3.6 shows how often a GLOTTALIZED vowel was selected by each participant for each type of stimulus (organized by gender and dialect of the speaker). We can see that stimuli with a GLOTTALIZED vowel tended to elicit more GLOTTALIZED vowel responses and that a stimulus with a HIGH TONE vowel tended to elicit fewer GLOTTALIZED vowel responses, regardless of gender. Additionally, if we look at how the phonetic dimensions of pitch and glottalization affect the participant's response, we get the same effects regardless of the gender of the participant; both males and females are more likely to select a GLOTTALIZED vowel if the stimulus has more glottalization, higher initial pitch, and a larger pitch span.[31]

[31] Though males tended to select more GLOTTALIZED vowels as initial pitch increased, the effect of initial pitch on the response of male participants was nonsignificant. This could be due to the small sample size (5 participants).

Table 3.6: Results for task 1 by gender of participant and dialect of speaker of stimulus (Santa Elena participants only)
Each cell shows the percentage of times a GLOTTALIZED *vowel was selected by males (m) and females (f) for each stimulus (organized by dialect of speaker).*

	stimulus = *k'a'an* (GLOTTALIZED vowel)											
	Mérida							Sisbicchén				
	1	2	3	4	5	6	7	8	9	10	11	12
f	78	89	100	100	100	100	100	33	33	67	78	44
m	40	80	60	80	80	60	100	0	60	80	40	100
	Santa Elena											
	13	14	15	16	17	18	19	20	21	22	23	24
f	100	44	33	60	78	89	67	89	44	56	22	22
m	80	60	40	22	100	40	60	40	20	40	20	40
	stimulus = *k'áan* (HIGH TONE vowel)											
	Mérida							Sisbicchén				
	1	2	3	4	5	6	7	8	9	10	11	12
f	67	22	22	44	44	22	33	33	11	44	33	44
m	40	20	60	40	40	60	60	60	20	0	40	40
	Santa Elena											
	13	14	15	16	17	18	19	20	21	22	23	24
f	44	22	22	33	56	44	44	22	33	44	33	33
m	40	0	60	40	100	60	40	60	20	0	40	0

The second hypothesis, that listeners use different strategies on the basis of gender of the speaker is also not substantiated. When the speaker is male, listeners are more likely to select a GLOTTALIZED vowel when the stimulus has more glottalization, higher initial pitch, and a larger pitch span. When the speaker is a female, the same is true, except that the effect of initial pitch is nonsignificant. This nonsignificant result could be taken as evidence in favor of the hypothesis, but as it is the only evidence it is rather weak. This study was not designed for the purposes of detecting gender-based differences, and so as we divide the data into more specific categories, the sample size decreases, making nonsignificant results more likely. Thus, I conclude that this data does not provide any strong evidence that perceptual strategies differ on the basis of the gender of the speaker or of the listener. However, some of the nonsignificant results may give other researchers reasons to explore this area further.

In summary, participants from Santa Elena were able to do just better than chance at identifying which word the speaker said. Listeners (of both genders) make use of all the cues that are distinguished in production. If a stimulus has more glottalization, higher initial pitch, or a larger pitch span, then listeners are more likely to select a GLOTTALIZED vowel.

3.3.3 Results for All Locations

In this section we will see how participants from other dialects reacted to task 1. Because of the small numbers of participants in each dialect group, the focus of this section (and of §3.4.3) is exploratory data analysis rather than hypothesis testing. Table 3.7 shows that most dialect groups selected more GLOTTALIZED vowels if the stimulus had a GLOTTALIZED vowel and more HIGH TONE vowels if the stimulus had a HIGH TONE vowel. The exception to this is Sisbicchén, where more GLOTTALIZED vowels were chosen for both stimuli.

Table 3.7: Results for task 1, percentage of times a GLOTTALIZED vowel was selected on the basis of stimulus vowel shape
(The number in parentheses after a town name indicates the number of participants from that town; the number in parentheses after a percentage indicates the number of tokens.)

	stimulus	
	k'a'an (G)	*k'áan* (H)
Sisbicchén (3)	63% (45)	56% (40)
Xocén (2)	69% (33)	35% (17)
Yax Che (2)	63% (30)	31% (15)
Mérida (2)	58% (28)	29% (14)
Santa Elena (14)	63% (212)	37% (124)

The effect of initial pitch and pitch span on the listeners' responses is shown in Tables 3.8 and 3.9. For the two towns in the west, as initial pitch increases, participants from Mérida and Santa Elena are more likely to select a glottalized vowel (except for some irregularities at both ends of the scale in Santa Elena). For Sisbicchén and Xocén, initial pitch does not have much of an effect on the listeners' responses. Listeners from Yax Che pattern with the listeners from the west in that there is some preference for GLOTTALIZED vowels to have higher initial pitch, but the initial pitch category of [8] triggers 0 GLOTTALIZED vowel responses. Recall that Yax Che patterned more with the eastern towns of Sisbicchén and Xocén in production, but there was one notable exception in that LOW TONE vowels tended to have higher pitch than HIGH TONE vowels in Yax Che only. Thus, in Mérida and Santa Elena, where both vowel shapes have a falling pitch contour but GLOTTALIZED vowels tend to have higher pitch initially, the response of listeners is influenced by initial pitch. In Sisbicchén and Xocén, where HIGH TONE and GLOTTALIZED vowels do not differ by pitch in production, listeners do not use pitch in perception. Finally, in Yax Che, where HIGH TONE vowels tend to have low pitch in production, listeners behave somewhat like those in the west and prefer GLOTTALIZED vowels to have higher pitch.

Table 3.8: Results for task 1, percentage of times a GLOTTALIZED vowel was selected on the basis of stimulus initial pitch
(The number in parentheses indicates the number of participants from that town.)

	initial pitch category					
	-2	0	2	4	6	8
Sisbicchén (3)	58	52	54	67	50	100
Xocén (2)	58	50	38	62	50	50
Yax Che (2)	50	29	35	65	75	0
eastern dialect (7)	56	45	44	65	57	57
Merida (2)	33	36	31	62	75	100
Santa Elena (14)	52	44	38	59	79	50
western dialect (16)	50	43	37	59	78	56

Table 3.9: Results for task 1, percentage of times a GLOTTALIZED vowel was selected on the basis of stimulus pitch span
(The number in parentheses indicates the number of participants from that town.)

	pitch span category					
	2	4	6	8	10	12
Sisbicchén (3)	64	53	50	73	67	67
Xocén (2)	32	53	50	60	67	88
Yax Che (2)	23	32	63	70	67	100
eastern dialect (7)	43	47	54	69	67	82
Merida (2)	18	29	50	80	67	100
Santa Elena (14)	33	41	59	63	71	84
western dialect (16)	31	40	58	65	71	86

All locations except Sisbicchén prefer for GLOTTALIZED vowels to have a larger pitch span. We expect this because the GLOTTALIZED vowel is produced with glottalization in all locations and glottalization is associated with a large pitch span.

Sisbicchén is also anomalous in that glottalization does not affect the responses of these listeners, as shown in Table 3.10, though a full glottal stop seems to trigger a response of a GLOTTALIZED vowel. All the data with respect to participants from Sisbicchén seems to suggest that they were simply guessing during the entire task. If this is true, it could be related to the fact that most of the stimuli were spoken by speakers of the western dialect. Given how different productions from the west (Santa Elena and Mérida) were from productions from the east (Sisbicchén) in production study 1, it is possible that listeners from Sisbicchén were thrown off and were unsure of how to respond.

Table 3.10: Results for task 1, percentage of times a GLOTTALIZED vowel was selected on the basis of stimulus glottalization type
(The number in parentheses indicates the number of participants from that town.)

	glottalization type			
	ng	wg	cr	gs
Sisbicchén (3)	59	56	58	83
Xocén (2)	40	63	100	75
Yax Che (2)	27	83	63	100
eastern dialect (7)	44	65	71	86
Merida (2)	32	71	38	75
Santa Elena (14)	39	68	55	93
western dialect (16)	38	69	53	91

For all other locations, a stimulus with modal voice was most likely to trigger a response of *k'áan* (with a HIGH TONE vowel). Stimuli with some form of glottalization were more likely to trigger a response of *k'a'an* (with a GLOTTALIZED vowel). This is expected as GLOTTALIZED vowels tend to be produced with glottalization whereas HIGH TONE vowels are not.

3.3.4 Conclusions from Task 1 (Natural Stimuli)

The results of task 1 show that most listeners are able to do better than chance at determining whether the speaker said a word with a HIGH TONE vowel or a GLOTTALIZED vowel, and that, for the most part, perception mirrors production. Participants from the western towns (Mérida and Santa Elena) use glottalization and pitch when choosing between these two vowel shapes, just as these two phonetic dimensions distinguish the vowel shapes in production. Participants from Sisbicchén and Xocén do not use initial pitch as a cue to perception, and this phonetic dimension does not distinguish the two vowel shapes in production. Participants from Yax Che do use initial pitch in perception. This is somewhat surprising given that GLOTTALIZED vowels did not regularly have higher initial pitch in the production data from these speakers, but it is also the location where HIGH TONE vowels tend to have lower pitch than LOW TONE vowels. Again, it is clear that more work is needed with speakers from the eastern towns in order to develop a better understanding of vowel shape in this dialect.

3.4 Task 2: Pitch and Glottalization as Cues to Distinguishing HIGH TONE and GLOTTALIZED Vowels in Manipulated Tokens

The results from task 1 indicate that perception does indeed mirror production; the phonetic dimensions that differ in the production data act as cues to perception. However, the stimuli of task 1 are not ideal for a full analysis of how different values for different phonetic dimensions effect perception. Because the stimuli in task 1 were tokens of natural speech, they contain all and only the phonetic values that those 24 speakers actually produced. For example, only one GLOTTALIZED vowel was produced with an initial pitch value between 1 and 2 s/b, and this stimulus was produced with modal voice. There is thus no evidence from task 1 as to how a participant might react to a stimulus with an initial pitch between 1 and 2 s/b and weak glottalization, creaky voice, or a glottal stop. In other words, the 48 natural stimuli display accidental gaps. A token with an initial pitch of 1.5 s/b and a glottal stop is perfectly legal (for a GLOTTALIZED vowel); it just was never produced for *k'a'an* in production study 1.

Another potential problem with the stimuli used in task 1 is that real Yucatec Maya speakers may be using cues to distinguish GLOTTALIZED from HIGH TONE vowels that we do not know about. Perhaps the onset and coda consonants contain information on vowel shape that is used by the listener. It could thus be the case that there are more factors that are affecting the listeners' choices than we are aware of. The second task was designed to use a more controlled setting to shed some light on how pitch and glottalization (and their interaction) influence the listener when discriminating between HIGH TONE and GLOTTALIZED vowels. In order to do this, stimuli were manipulated to fit every combination of four different glottalization types and four different initial pitch values. We will see, however, that participants did not react to manipulated stimuli in the same way as they did to natural stimuli, and these differences will be discussed in §3.5.

3.4.1 Procedures and Manipulation of Stimuli

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'áa̰n] 'strong' and *cháak* [tʃáak] 'rain' vs. *cha'ak* [tʃáa̰k] 'starch'.[32] Stimuli were manipulated from one production of *k'an* 'ripe' and *chak* 'red' (both

[32] As discussed in §3.3.1, the minimal pair *k'a'an* and *k'áan* is robust. 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 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 this language.)

are current lexical items with a SHORT vowel) as spoken by a male from Mérida who did not participate in the perception experiments. Each stimulus immediately followed the frame sentence *Tin wa'alaj* ___ 'I said ___', which was an unmanipulated recording spoken by the same male. A frame sentence was used in this task because I was worried 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. In this particular frame sentence, the stimulus is phrase-final, and, as explained in Chapter 2, GLOTTALIZED and HIGH TONE vowels are produced similarly in phrase-final and isolated contexts.[33]

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 was 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 in the west).

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.

[33] 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, and so there is no reason to expect it to affect the pronunciation of the target word.

3.4.1.1 Manipulation of Pitch

Four different pitch contours were created using PRAAT and were defined by three points – the beginning of the vowel (analogous to time point 1, as used to define pitch contours in Chapter 2), 75 ms after the beginning of the vowel (between time points 2 and 3), and the end of the vowel (analogous to time point 5). 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*). Because the speaker of the stimuli participated in production study 2, we know what his baseline is in sentence-final context: it is 125.77 Hz. Thus, the pitch contours used in task 2 begin with -0.11, 1.86, 3.62, and 5.22 s/b; these pitch values will henceforth be abbreviated as L (low), ML (mid-low), MH (mid-high), and H (high).[34] The resulting contours are illustrated in Fig. 3.3, where (a) shows the four pitch tiers that were created by hand in PRAAT and (b) shows the actual pitch contours of the resulting stimuli after publishing the resynthesized token (for *cháak*/*cha'ak* with modal voice). In Fig. 3.4, we see where these manipulated contours fall relative to this speaker's natural average productions of these vowel shapes.

[34] 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 I have already made the claim the 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.

Figure 3.3: Pitch contours for stimuli used in task 2

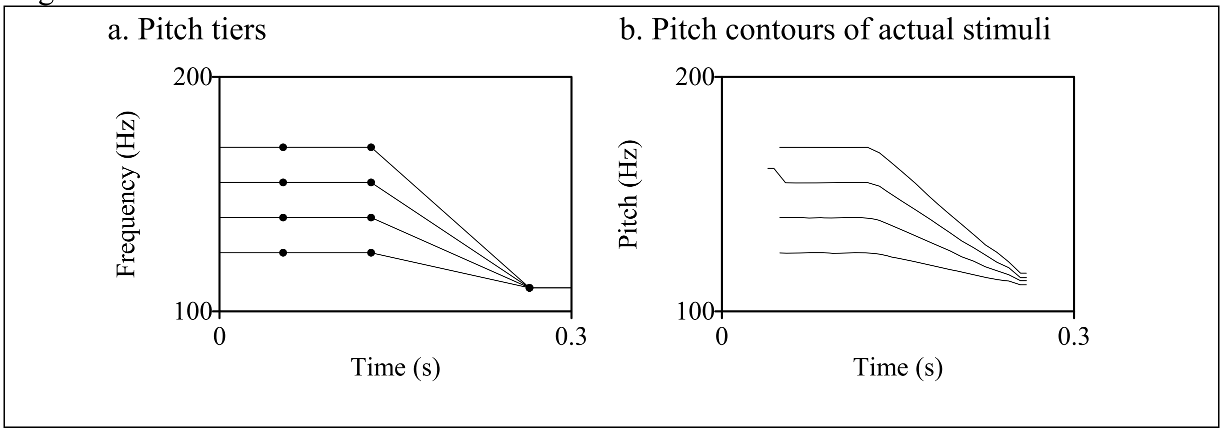
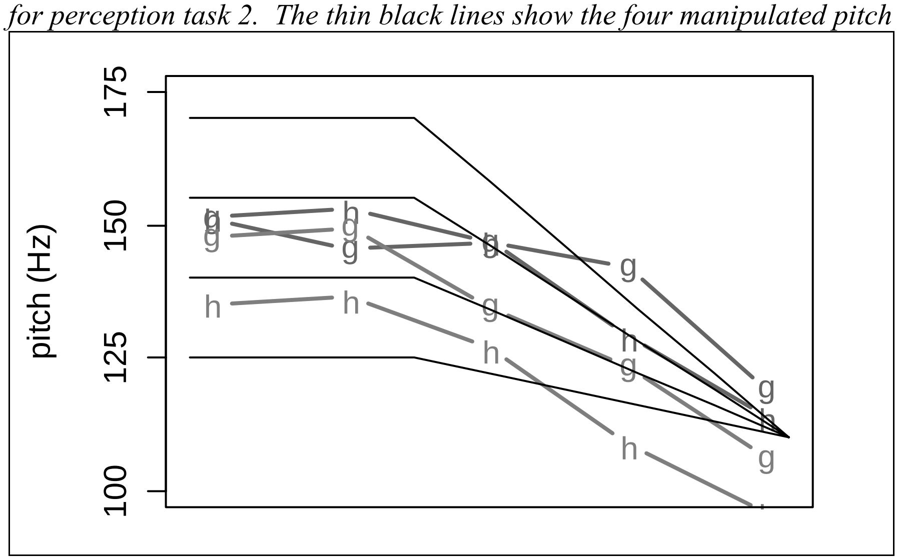


Figure 3.4: Comparison of manipulated pitch contours with speaker's natural productions

The thick gray lines show the average pitch contours for HIGH TONE *'h' and* GLOTTALIZED *'g' vowels (from production study 1 (lighter gray, lower pitch values) and production study 2 (darker gray, higher pitch values)) 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.*



Because the lowest pitch value for each of these four contours is the same (110 Hz), the pitch span of each contour is perfectly correlated with initial pitch. The pitch spans for each contour (in semitones) are 2.21 (for L), 4.18 (for ML), 5.94 (for MH), and 7.54 (for H).

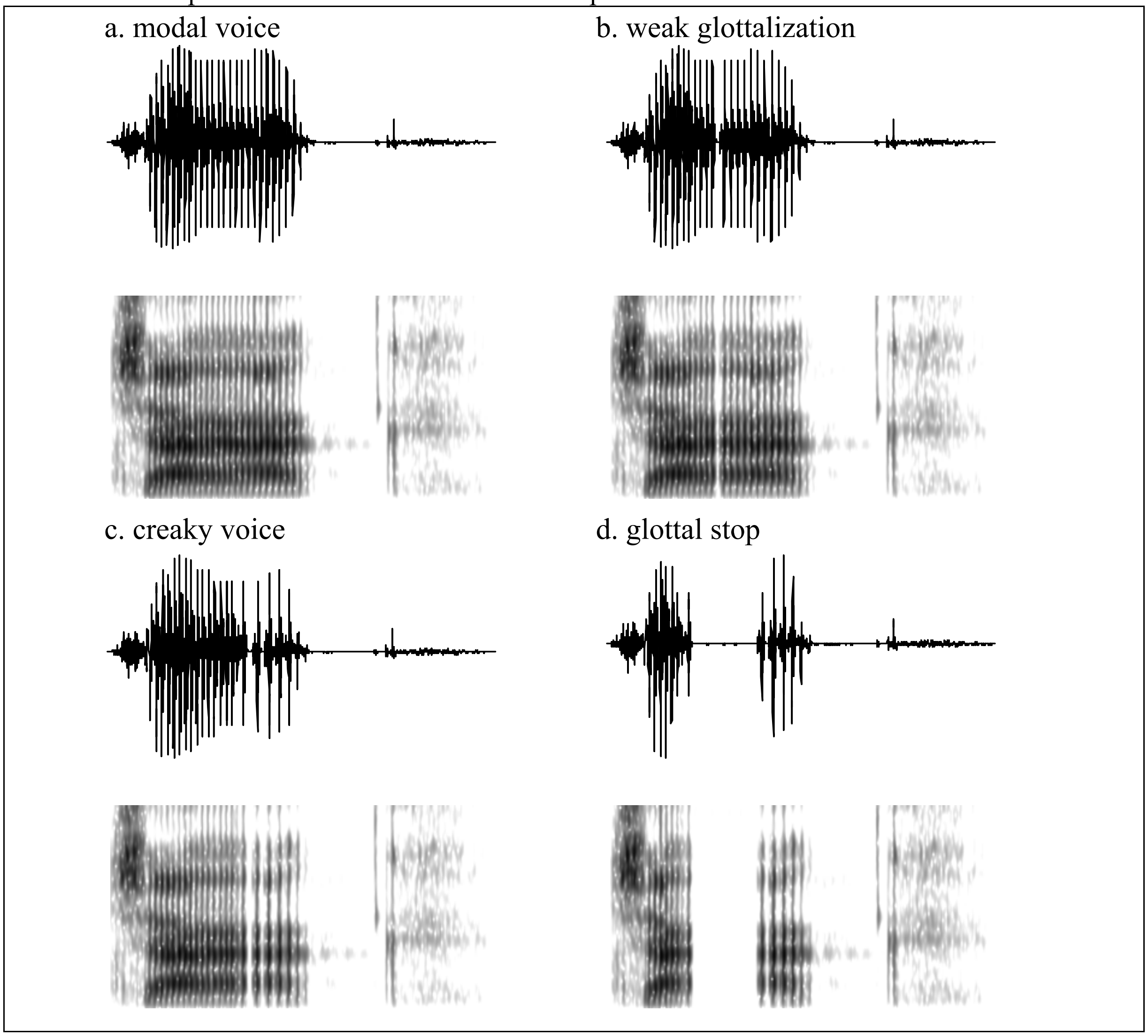
3.4.1.2 Manipulation of Glottalization

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. 3.3a) 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 the stimulus is synthesized with such a pitch contour, the resulting stimulus contains a very brief portion of what sounds like creaky voice.

Tokens with "creaky voice" were manipulated by starting with the weak glottalization pitch tier (the one with one 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. As discussed in §2.1.3.1, creaky voice is consistently produced with lower intensity than modal voice. 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 the longer duration of creaky voice were used as a starting point. About 75 ms was 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 vowels with a glottal stop are 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 showing the four different glottalization types are displayed in Figure 3.5.

Figure 3.5: Spectrograms and waveforms for glottalization types used in task 2
These stimuli represent *cháak*/*cha'ak* with an initial pitch of 140 Hz.



3.4.2 Results for Santa Elena

For this task, each manipulated stimulus is essentially ambiguous; there is no intended target of a GLOTTALIZED or HIGH TONE vowel. The stimuli are identical except for pitch and glottalization, and so these are the only factors that can influence the listeners' decisions. The results for listeners from Santa Elena are displayed in Table 3.11. Here we see that, in task 2, the participants are making their decisions on the basis of glottalization alone. Initial pitch has no effect on whether a GLOTTALIZED or HIGH TONE vowel is perceived.

Table 3.11: Percentage of times participants selected a word with a GLOTTALIZED vowel (Santa Elena only)

	initial pitch category			
glottalization type	L (-0.1)	ML (1.9)	MH (3.6)	H (5.2)
glottal stop	79%	85%	77%	83%
creaky voice	61%	63%	63%	69%
weak glottalization	44%	44%	41%	37%
modal	27%	25%	23%	29%

significant effect of glottalization: Wald $\chi^2(3)=35.8$, $p < .0001$
nonsignificant effect of pitch: Wald $\chi^2(3)=4.6$, $p =.20$
nonsignificant interaction: Wald $\chi^2(9)=11.2$, $p =.26$
(standard errors are adjusted for multiple observations within subjects)

This result is surprising given that pitch distinguishes the two vowel shapes in production and that listeners use this cue in task 1. Why do they now listen for glottalization alone? The most obvious answer is that the manipulated stimuli elicited different behavior than the natural stimuli did. The question then is whether or not this different behavior is a function of the grammar (and if so how). The data for each participant from Santa Elena (see Table 3.13) shows that they all behaved in the same way. There is a nonsignificant effect of pitch for each participant. For all but three participants, there is a significant effect of glottalization. This means that participants were not just randomly guessing (or we would expect to see different behavior for each participant). Were their responses based on some sort of adjusted language-specific perception grammar or were their responses based on some biological (and hence language-universal) bias? I argue that these responses are determined by a language-specific grammar in §3.5.

3.4.3 Results for All Locations

When we look at all locations (Table 3.12), we see that listeners from Mérida behaved exactly like those from Santa Elena. Glottalization alone and not pitch is correlated with the participant's response. Thus, in all production and perception tasks, the participants from the western towns of Mérida and Santa Elena show mostly the same patterns.

Surprisingly, pitch seems to affect the responses of participants from the eastern towns, but the patterns are not robust. Depending on glottalization category and location, different pitch values trigger more GLOTTALIZED vowel responses. For example, in Sisbicchén in the creaky voice category, the pitch categories of L and H trigger more GLOTTALIZED vowel responses, but in Xocén in the creaky voice category, the ML pitch category triggers the most GLOTTALIZED vowel responses. We cannot generalize about the effect of pitch on a participant's selection in the eastern towns, though we can conclude that listeners in all locations react differently to natural versus manipulated stimuli.

Table 3.12: Percentage of times participants selected a word with a GLOTTALIZED vowel (all locations)

		initial pitch category			
	glottalization type	L (-0.1)	ML (1.9)	MH (3.6)	H (5.2)
Sisbicchén	glottal stop	89%	94%	89%	100%
	creaky voice	83%	72%	67%	83%
	weak glot.	50%	56%	67%	78%
	modal	17%	28%	11%	6%
Xocén	glottal stop	92%	83%	100%	100%
	creaky voice	67%	83%	67%	75%
	weak glot.	25%	58%	25%	58%
	modal	8%	17%	0%	8%
Yax Che	glottal stop	83%	100%	92%	92%
	creaky voice	75%	83%	75%	83%
	weak glot.	67%	67%	83%	92%
	modal	17%	8%	17%	5%
Mérida	glottal stop	92%	67%	75%	75%
	creaky voice	75%	58%	100%	58%
	weak glot.	42%	33%	42%	75%
	modal	25%	33%	25%	25%
Santa Elena	glottal stop	79%	85%	77%	83%
	creaky voice	61%	63%	63%	69%
	weak glot.	44%	44%	41%	37%
	modal	27%	25%	23%	29%

3.5 Comparing the Results of Task 1 and Task 2: The Quality of the Linguistic Situation Influences the Grammar

We have seen that participants from the west used multiple cues in discriminating between HIGH TONE and GLOTTALIZED vowels when listening to natural stimuli but that these same participants attend to glottalization alone when listening to manipulated stimuli. Because the stimuli of task 1 were tokens of natural speech, it seems safe to assume that the participants were using the same perceptual strategies that they would use in regular linguistic situations. The manipulated stimuli of task 2, on the other hand, may be considered to constitute a 'degraded' linguistic situation. The effect of stimulus quality on perception has been documented in the literature; van Hessen and Schouten (1999) discuss how stimulus quality affects "categorical perception", which, for them, equates with a situation where listeners can only distinguish between phonemic categories and cannot distinguish among (phonetically different) members of a phonemic category (following Liberman et al. 1957). This is the only situation that is defined as true categorical perception, and so they discuss degrees of categoricalness with respect to how well the strict criteria are met. Their conclusion is that there is an increase in categorical perception as stimulus quality increases.

This theory in and of itself is neither contradicted nor confirmed by the results of task 1 and task 2 because neither results showed categorical patterns. I do not even know how something like categorical perception could be possible with the GLOTTALIZED and HIGH TONE vowels of Yucatec Maya because most phonetic forms that are legal productions of one underlying category are also legal productions of the other underlying category. However, when van Hessen and Schouten (1999: 58) interpret their results, they make a claim that is highly relevant: "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 task 1 and task 2, it seems clear that participants in task 1 (with natural stimuli) "had to listen to the full spectrum and all its subtle interacting cues", while participants in task 2 did not. So why did the participants in this task 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?

We can rule out the last possibility by looking at the data for each participant separately. As shown in Table. 3.13, all participants from Santa Elena behaved in a similar fashion in task 2. For all but three participants, there is a significant effect of glottalization type on response and for all participants there is a nonsignificant effect of pitch.

Table 3.13: Percentage of times each participant selected a word with a GLOTTALIZED vowel (Santa Elena only, task 2, manipulated stimuli)

Each outlined block of data comes from one participant; the shaded cells denote those participants for which there was not a significant effect of glottalization type.

	L	ML	MH	H	L	ML	MH	H	L	ML	MH	H
mod.	17	17	0	0	0	0	0	0	17	33	0	67
wg.	17	50	67	67	50	17	50	33	50	50	50	17
cr.	100	83	100	83	50	100	100	100	83	50	50	67
gs.	100	100	100	100	100	100	100	100	83	100	100	100
mod.	0	0	17	0	50	33	33	17	67	67	67	67
wg.	33	17	0	17	33	50	83	67	17	60	67	33
cr.	100	83	33	67	83	50	50	67	33	33	33	50
gs.	67	100	100	100	67	67	67	100	33	50	50	67
mod.	50	50	33	17	33	33	50	33	17	0	17	33
wg.	50	50	33	33	67	67	50	17	67	50	17	50
cr.	33	50	67	67	83	100	83	83	67	67	67	67
gs.	67	100	50	50	100	100	100	67	83	100	67	100
mod.	33	33	33	67	0	33	0	0	17	17	17	17
wg.	0	0	33	33	33	33	0	0	67	33	17	33
cr.	0	33	0	67	67	67	100	67	17	17	33	50
gs.	33	0	67	33	100	100	100	67	67	50	50	67
mod.	33	33	33	33	33	0	0	100				
wg.	50	50	33	67	67	100	67	33				
cr.	33	50	67	50	100	100	100	100				
gs.	100	100	50	83	100	100	100	100				

The fact that almost all participants behaved in the same way suggests that they were not simply guessing but instead were making some sort of informed decision. The question that remains is whether or not the perceptual behavior exhibited in task 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. 3.4). Further experimentation would need to be done in order to determine how non-Yucatec Maya speakers might react to the stimuli of task 2, but in the absence of such experimentation I find these "universal" explanations unsatisfying. Furthermore, the patterns of production in Yucatec Maya 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 3.14 we see the percentage of times each phonetic value (used with the stimuli of task 2) was produced as a GLOTTALIZED vowel (as compared to how often it was produced as a HIGH TONE vowel) in the data set of 685 tokens that is used for the learning simulations of Chapter 5 (see §5.1). This table shows that, for example, of all the tokens that contained modal voice 36% of them were productions of GLOTTALIZED vowels (meaning 64% of them were productions of HIGH TONE vowels). 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 make 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 (64% 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 [0] or [2] as a HIGH TONE vowel and all tokens with an initial pitch of [4] or [6] as a GLOTTALIZED vowel, they will only be right about half the time. It 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 Yucatec Maya.

Table 3.14: Percentage of times each phonetic category is produced as a GLOTTALIZED vowel (out of how many times it is produced as either a HIGH TONE or GLOTTALIZED vowel)

glottalization type			
mod	wg	cr	gs
36	93	95	100
initial pitch (s/b)			
0	2	4	6
36	46	50	63
pitch span (semitones)			
2	4	6	8
36	42	49	65

In Yucatec Maya, 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 task 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.

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 dissertation 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 Yucatec Maya 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.

One final point is relevant on this topic. The above discussion is relevant to the notion of "minimal contrast". I believe the results of task 2 present evidence that HIGH TONE and GLOTTALIZED vowels are *minimally contrastive* on the basis of glottalization. This idea was at least implied in Chapter 2, where I proposed that the phonological forms of these vowels are /v́v/ and /v́v̰/, which differ only by glottalization. There are various proposals as to what minimal contrast is and how it does (or doesn't) affect the grammar. For example, Campos-Astorkiza (2007) expresses the view that minimal contrast plays a vital grammatical role, while Dresher (in press) expresses the view that there is no notion of minimal contrast. This is not the place to fully evaluate these claims, but the data obtained through both production and perception experimentations points to glottalization as the most important phonetic cue in signaling the contrast between HIGH TONE and GLOTTALIZED vowels.[35] I believe that it is appropriate to claim that these two vowel shapes minimally contrast on the basis of glottalization. The results of task 2 suggest that listeners not only have knowledge of this type of minimal contrast, but use this knowledge to alter their grammar in degraded language situations.

3.6 Chapter Summary

For all locations, the results of task 1 differ from the results of task 2. Participants from the western towns were influenced by both pitch and glottalization in task 1 but by glottalization alone in task 2. It is harder to generalize about participants from the eastern towns because of the smaller sample sizes. Participants from Sisbicchén appeared to be randomly guessing in task 1. While both pitch and glottalization influenced the responses of these participants in task 2, only the pattern with regard to glottalization is consistent in that more glottalization elicited more GLOTTALIZED vowel responses. Participants from Xocén and Yax Che primarily used the cue of glottalization in both task 1 and task 2, though there are also some (sporadic) effects of pitch for these listeners.

Because the stimuli for task 1 were completely natural, it is most likely that participants reacted to this task as they would in a real language setting, whereas the manipulated stimuli of task 2 brought out different strategies for perception. Based on this assumption, I will use only the results of task 1 (from the Santa Elena participants) for analysis in Chapter 5. These results indicated that listeners use all the cues available to them

[35] One of Dresher's (in press: §8.7) criticisms of the notion of minimal contrast in phonology is that there are always multiple phonetic cues that can be used to distinguish two phonemes. This is of course true of the contrast between HIGH TONE vowels and GLOTTALIZED vowels in Yucatec Maya, but what the data shows us is that one cue is the most useful. Thus I do not claim that the two vowel shapes are minimally contrastive in that they are differentiated by one and only one phonetic dimension, but that they are minimally contrastive in that one and only one phonetic dimension is the best indicator of the contrast.

in performing the discrimination task. Because glottalization, initial pitch, and pitch span are all phonetic dimensions that distinguish HIGH TONE from GLOTTALIZED vowels in production, listeners attend to all of these cues in perception (in the western dialect). In Chapter 5, the Bidirectional Model will be used to account for this correlation between the production and perception grammar.

With regard to task 2, listeners focused on glottalization alone, and this was likely a reaction to stimulus quality. I proposed in §3.5 that listeners adapted their perception grammar in the face of a degraded language setting. Because glottalization is the cue that is mostly likely to lead to successful comprehension on its own, this is the cue that listeners attended to.

4

THEORETICAL ANALYSIS AT THE PHONETICS-PHONOLOGY INTERFACE

It is now well understood that minute details of speech production that would not usually be considered phonological (i.e. categorical or contrastive) are subject to language-specific restrictions (as compared to the older view expressed in Chomsky and Halle (1968) that anything not dictated by the phonology is subject to universal laws; see also §1.1.1). Consider, for example, the English word *sad* /sæd/. We know that the jaw is lower for the production of [s] than it would be in the word *seed* /sid/ (Amerman et al. 1970) and that the vowel is longer than it would be in the word *sat* /sæt/ (Chen 1970). Furthermore, just how low the jaw is and just how long the vowel is are language-specific features of speech production.

It is clear that the sound grammar must not only account for contrast and categorical allophonic variation but also for these fine-grained details of speech production. Over the past few decades, there have been many proposals for how the grammar should handle the continuous phonetic aspects of speech. Boersma (2007a) identifies three main types of models that account for phonetics and phonology.[36] In a *serial* model, an underlying form is first mapped onto a phonological surface form, which is then mapped onto a phonetic form. In a *two-level* model, there is only one mapping – the underlying form is directly mapped

[36] Boersma (2007a: 2025) actually lists five types of models, but three of them of are variations of his own model.

onto a phonetic form. In Boersma's own model, a *parallel* model, the production grammar simultaneously maps an underlying form onto both a phonological surface form and two different phonetic forms. Each of these types of models is discussed in more detail in this chapter. I use Boersma's parallel model (Bidirectional StOT) for analysis in this paper, and I believe Boersma's theoretical arguments against the serial and two-level models are sound. It will also be worthwhile, though, to discuss some of the insights of versions of serial and two-level models.

In this chapter I first discuss some of the methods used to distinguish phonetic processes from phonological processes in §4.1. I then summarize two models of analysis for the phonetics-phonology interface: Keating's (1990) Window Model (a serial model, §4.2.1) and Flemming's (2001) Unified Model (a two-level model, §4.2.2). In §4.3 I present the model of analysis used in this paper – Boersma's (2007a) Bidirectional Model; a chapter summary follows in §4.4.

4.1 Distinguishing Phonetics from Phonology

Much literature of the last three decades assumes a distinction between two sound grammars (see, for example, Pierrehumbert 1980; Fourakis and Port 1986; Keating 1985a, 1990; Cohn 1990, 1993, 1998; Zsiga 1997; as well as *Journal of Phonetics* 18 (1990), which is a special issue on the phonetics-phonology interface). In these papers, a variety of models are proposed with one thing in common: a distinction between phonetics and phonology on (at least) the basis of gradience. Cohn (1998: 25) summarizes: "phonology is the domain of the discrete, whereas phonetics is the domain of the continuous". She then states that phonological representations capture abstract patterns, while phonetic representations show a physical realization. Finally, Cohn presents the adopted standard (following Keating 1985a) that phonological rules feed into phonetic rules. These classic views about the phonetics-phonology interface are summarized in (4.1) (see also Cohn (2006) for a current review of the state of research in phonetics in phonology).

(4.1) common viewpoints regarding a distinct phonetics and phonology

	phonetics	phonology
domain	gradient processes	categorical processes
type of representation	specific in space and time	abstract symbols
order of application	second	first

In more recent literature, the two-module approach, as outlined above, has been challenged. Boersma (1999b) argues that the criteria used to distinguish phonetics and phonology fail on several counts. For example, the idea that phonetic rules are somewhat optional and at least paradigmatically conditioned is also true of some postlexical phonological rules. He cites nasal stop assimilation to following oral stops in Dutch as an example of an optional phonological process: assimilation rarely happens in "clear Dutch" though it often happens in "normal Dutch". Boersma's conclusion is that phonology and phonetics are included in a single module. This is the same conclusion reached by Flemming (2001), who argues that phonetic and phonological processes are motivated by the same factors and that distinct modules fail to account for these similarities. For example, coarticulation (phonetic) and neutralization (phonological) are both motivated by the desire for ease of articulation and should thus be accounted for in the same grammar that utilizes this motivation. Others who support a unified account of phonetics and phonology include Kirchner (1997) and Byrd (1996).

Though both Flemming and Boersma dispute a complete separation of phonetics and phonology, they differ in degree of integration. Boersma argues for a model that uses both phonological (discrete) and phonetic (continuous) representations, whereas Flemming argues for a two-level model such that the only output of the production grammar is a continuous phonetic representation. These three different proposals for how to incorporate phonetics into the overall sound grammar are summarized in Table 4.1.[37]

Table 4.1: Proposals for a production grammar of phonetics and phonology

modules	output representations	details	example source
2	2	a phonological module generates a discrete abstract output which feeds into a phonetic module that generates a continuous output: |underlying form| → /surface form/ → [phonetic form]	Keating 1990
1	2	discrete phonological outputs and continuous phonetic outputs are generated simultaneously: |underlying form| → /surface form/, [phonetic form]	Boersma 2007a
1	1	only a continuous phonetic output is generated by the grammar: |underlying form| → [phonetic form]	Flemming 2001

4.2 Significant Advancements in Modeling the Phonetics-Phonology Interface

4.2.1 Keating's Window Model

In this oft cited work (Keating 1990, henceforth simply Keating), Keating claims that the phonetic component of the grammar operates on the output of the phonology, which is a symbolic representation. The phonology is discrete and makes use of binary features, and Keating stresses that this makes phonological representations idealizations. Temporal relationships are denoted by segmentation, which indicates sequencing though the segments themselves are otherwise timeless. The labeling of segments denotes categorical actions of the articulators. For example, the label /z/, as opposed to the label /d/, tells us that the tongue approaches the alveolar ridge (instead of making a full closure with the alveolar ridge), but it does not tell us exactly how close the tongue and alveolar ridge are nor how long the tongue

[37] I follow Boersma (2007a) in using a vertical line to denote underlying forms, slashes to denote phonological (discrete) surface forms, and brackets to denote phonetic (continuous) forms (see §4.3).

remains in that position.

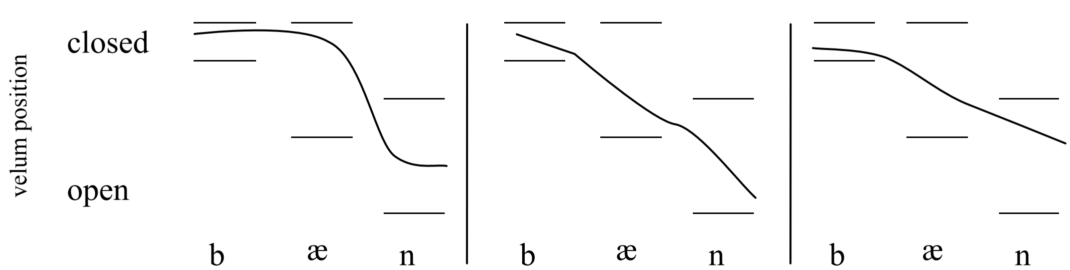
The job of the phonetic module is to map a symbolic representation onto a physical representation "which, like speech, exists in continuous time and space" (p. 451). The primary difference between phonetics and phonology is thus that phonetic representations are explicit with respect to temporal and spatial structure, whereas phonological representations are only vague indicators of such structure. Keating does mention a commonly stated distinction between phonetics and phonology as one of category, with phonetics being more gradient and phonology more categorical, but she is rightly cautious in warning that there are degrees of gradience such that this feature alone cannot distinguish phonetics from phonology.

Assuming that phonetics is the component of the grammar that takes symbolic representations of sound and maps them onto explicit models of space and time, Keating proposes that this is accomplished through the use of *windows*. A window defines an acceptable range of articulator positions (e.g. velum position, jaw height, lip position, frequency of vocal fold vibration, etc.) with respect to a particular phonological feature. For example, for the vowel feature [+high], there is a window that defines the range of tongue (and jaw, etc.) positions that are acceptable for [+high] sounds. Keating explains that a window is not a range around a mean. In fact, the mean is irrelevant in her system – minimum and maximum values are what define the range of the window. In this case, it would not matter what the average tongue position is in the production of high vowels, it would only matter what the minimum and maximum tongue heights are in such productions.

The use of windows captures an important intuition about the variability of phonetic forms. At the phonological level, a high vowel will always bear the feature [+high], but at the phonetic level the same high vowel may be produced with different F_1 values. All of these F_1 values may be grammatical, and so the phonetic grammar must be able to associate a feature like [+high] with *a range of F_1 values*, and not just with single grammatical F_1 value. Because windows define a range of values, they can account for the fact that those values vary from utterance to utterance.

In order to illustrate how the window model works, we will look at a well known phonetic phenomenon in English. Cohn (1993) proposes that vowel nasalization in English occurs at the phonetic level, and so we can assume that the phonological output for a word like *ban* is simply /bæn/, without vowel nasalization. According to the window model, different windows for velum position are stipulated by the grammar for the [-nas] consonant [b], the [-nas] vowel [æ], and the [+nas] consonant [n].[38] The actual velum positions for the duration of the word are then derived by drawing lines that pass through these windows. In (4.2), I show three different possibilities for how the velum position changes with time (each in accordance the same window specifications) during the production of *ban*.

(4.2) hypothetical window model analysis for velum position in /bæn/



In all the paths in (4.2), the vowel ends up with more nasalization than it started with, but it is not categorically "nasalized", which accurately represents the facts about vowel

[38] I assume that non-nasal consonants and vowels would be associated with different windows in English because nasalization is contrastive with consonants but not with vowels.

nasalization in English. In this way, the Window Model successfully accounts for variation, which is naturally present in phonetic data. Furthermore, the width of the window can constrain that variation in meaningful ways, such as the way the narrow window for velum position for /b/ does not allow the velum to open far enough for nasalization to occur and also limits how much nasalization the vowel can start with.

Despite the prevalence of citations of the Window Model in the literature, there is not much in the way of actual applications of the Window Model to language data. For this reason there are many basic questions about how to implement the model that have not been addressed. For example, how are the widths of the windows defined? Though Keating claims minimum and maximum values (as obtained through phonetic experimentation) should delimit the window, this seems too extreme. Certainly some minimum and maximum values should be considered speech errors, i.e. the speaker was shooting for a value near the edge of a window and missed. Are all paths through consecutive windows permissible or should we develop further limitations? With the windows in (4.2), it would be possible to draw a path through all three windows where the velum closes a bit during production of the vowel and then opens again. This is probably something that we want the grammar to forbid, and this could easily be done by prohibiting paths that take unnecessary turns. The key is, though, that we want our phonetic grammar to still be able to account for variation and also to constrain that variation in language-specific ways. While is seems that both of these goals could be accomplished by the Window Model, there is no literature showing how they are accomplished.

Boersma (2007a) presents arguments against the general class of serial models, of which the Window Model is one. In this paper, he analyzes the *h*-aspiré forms in French.

When the feminine indefinite article |ynə| precedes a vowel initial noun, the schwa is normally deleted, e.g. |ynə#ide| → [ynide] 'an idea'. However, schwa drop does not happen with the '*h*-aspiré initial' nouns, e.g. |ynə#ʔos| → [ynəos] 'a rise'. What is interesting here is that the schwa was retained even though the glottal stop was deleted, making the noun vowel-initial (which is the context where schwa drop normally occurs). Boersma argues that the cue constraints (which relate phonological surface forms to phonetic auditory forms) are ranked in such a way in this language that a schwa [ə] is perceived in the same manner as a 'creaky pause' and is hence mapped onto a |ʔ| in the process of comprehension. Thus it is cue constraints that allow the [ə] to be so interpreted (which would be part of the phonetic component of the grammar in a serial model) but the decision to delete the schwa or not is a discrete choice, and hence a phonological choice. In a serial model, there would be no way for the phonetic component of the grammar to tell the phonological component to retain the schwa. He says:

> "…the surfacing of the schwa in the form [ynəos] … was partly determined by phonetic considerations of perceptibility, considerations that in the serial model must have their locus in the phonetic implementation module. However, the presence or absence of schwa is a discrete choice that the speaker makes, and all serial models … would therefore very likely regard this choice as a discrete phonological decision. … However, if the separability of phonology and phonetics is taken seriously, this discrete phonological decision has no explanation in the module where it occurs, because the explanation is in a module that serially follows it and that it cannot see." (Boersma 2007a: 2026)

This is thus a case of phonetics influencing phonology. If phonetics is a separate module that follows phonology, as claimed by serial models, then there is no way for phonetics to influence phonology. Because there is language data that shows that such an influence exists, the serial model cannot be accurate.

Because of the theoretical arguments against serial models, I do not feel it would advantageous to attempt to answer some of the specific questions raised above with regard to implementing the Window Model. As will be discussed more in §4.3, the Window Model is an important precursor for current models of the phonetics-phonology, including Boersma's model, and so it deserves to be addressed in any review of phonetic grammar models.

4.2.2 Flemming's Unified Model

Flemming (2001, henceforth simply Flemming) challenges all two module approaches, like the window model. He claims that a two module approach is incapable of accounting for the fact that phonetic and phonological processes often have the same motivation. As discussed above, a similar critique is given by Boersma (2007a), who says that phonological processes can have phonetic motivation, and this is inexplicable in a model where the phonetics happens after the phonology.

Flemming's primary example is that of coarticulation (phonetic) and neutralization (phonological): in English, the /u/ in /tut/ is fronter than normal due to coarticulation with the coronal consonants, whereas in Cantonese this word is ungrammatical – the same input would map onto /tyt/, where a front vowel replaces a back one due to assimilation to the coronal consonants. If this aspect of English is analyzed in the phonetics and this aspect of Cantonese is analyzed in the phonology, how do we account for the fact that the processes have the same motivating factor? Because many phenomena show this parallel (and because other criteria for distinguishing phonetics and phonology, such as gradience, are problematic), Flemming claims that all aspects of sound are best analyzed with only one module of the grammar.

Flemming proposes a model that uses a constraint-based system, but differs from classic OT in two important details. First, constraints are finely detailed and refer to acoustic properties. For example, where OT might use a constraint *VOICEDOBS (voiced obstruents are banned), Flemming's model would use a constraint that bans some quantified value of the effort required to produce some articulatory correlate of a voiced obstruent. Furthermore, constraints are weighted instead of ranked.[39] This is a necessity given the types of constraints Flemming uses. Some constraints penalize any deviation from a specific target and other constraints penalize any articulatory effort. Neither of these constraints can ever completely dominate the other, i.e. speech is never produced without effort or by hitting every target exactly.

Using coarticulation/neutralization of vowels in the context of coronal consonants as an example, Flemming provides the constraints in Table 4.2 as examples of the types of constraints used by his model. The faithfulness constraints of OT are retained, but as explained above they now regulate specific acoustic properties. We see that IDENT($F2_C$) penalizes deviations from the target F_2 for a particular consonant. This means that a consonant α in the input is associated with a target F_2. The output candidates are fully specified in terms of acoustic features. An F_2 value in a phonetic output is compared to the target F_2 (for a given input) and a violation cost is assigned that depends on the difference between the F_2 of the output and the target F_2. This cost is multiplied by the weight of the constraint. The MINIMIZEEFFORT constraint can be considered a type of markedness constraint. According to the MINEFF constraint in Table 4.2, a consistent F_2 should be produced from the release of a consonant and into the vowel. The idea behind this constraint is that F_2 changes are correlated with tongue movement, which requires effort. A consistent F_2 means that no effort was exerted to move the tongue.

[39] When constraints are weighted, the optimal candidate is the one with smallest sum of violations.

Table 4.2: Constraints used in a unified model of phonetics and phonology (Flemming 2001)

	mathematical definition	verbal definition	cost of violation
IDENT($F2_C$)	$F2_C = F2_{target}$	A consonant α is produced with an F2 identical to the target F2 for α.	$w(F2_C - F2_{target})^2$
MINIMIZE EFFORT	$F2_C = F2_V$	F2 at the release of a consonant is identical to F2 in the middle of a following vowel.	$w(F2_C - F2_V)^2$
MINDIST = Δ	$\lvert F2_\alpha - F2_\beta \rvert \geq \Delta$	If α and β are contrasting segments, their F2s must differ by at least Δ.	$w(\lvert F2_\alpha - F2_\beta \rvert - \Delta)^2$ for $\lvert F2_\alpha - F2_\beta \rvert < \Delta$
MAXIMIZE CONTRASTS		Have contrasts.	$-w$

The last two constraints in Table 4.2 are unique to Flemming's model in that they have no straightforward correlates in classic OT. However these constraints have similar formulations in the literature on contrast (e.g. Padgett 1997, Barnes 2006, Campos-Astorkiza 2007). The MINDIST constraints define how different two sounds need to be in order to contrast. In Flemming's example, the /u/ of /tut/ is more like a /y/ than it would be in other contexts. The MINDIST constraint establishes whether or not the /u/ of /tut/ is still different enough from the /y/ of /tyt/ to maintain a contrast. The MAXCONTRAST constraint, of course, says to have contrasts. It assigns a negative violation (i.e. it decreases the sum of all violations for a candidate) for each contrast. Again, these constraints are in conflict. The more contrasts that are maintained, the less distinct the contrasting sounds will be (which is a necessary consequence given a vocal tract of finite size).

None of these constraints are ranked with respect to each other. Instead, their relative importance is determined by a weight. For example, if we consider just the IDENT and MINEFF constraints, a language can determine that a consonant should be produced with an F_2 closer to its target (by associating a higher weight with the IDENT constraint) or that a consonant should be produced with an F_2 closer to the F_2 of the following vowel (by associating a higher weight with the MINEFF constraint). It is clear that Flemming's system necessitates weighting as opposed to strict dominance of constraints. For example, even if the IDENT constraint has a really large weight, the MINEFF constraint is still able to exert some influence and thus account for minimal coarticulation.

Again, Boersma (2007a) provides general theoretical arguments against two-level models in general. According to Boersma, there is perceptual evidence for an intermediate phonological form. The process of perception is language-specific and so it must be constrained by a grammar. For example, as discussed in §1.1.1, given that voiceless stops in English have a larger VOT than voiceless stops in Spanish, we may expect English listeners to interpret a lower VOT value as a cue to a voiced stop, while a Spanish listener may interpret the same stimulus as a voiceless stop (which is exactly how English and Spanish listeners behave). Boersma argues that this mapping, called *prelexical perception*, happens before word identification. This means that it is not the mapping of a phonetic form onto an underlying form, but instead is the mapping of a phonetic form onto an intermediate phonological form, which he calls the surface form (see next section). This form is necessary to account for language-specific perception, and Boersma shows that it is also useful in the production grammar, thus invalidating a model with only two representations (underlying and phonetic).

Furthermore, Flemming's model is not designed to account for intra-language variation. In the unified model, constraints have consistent weights for a given language, and hence the same input will always map onto the same phonetic output. This means that if the grammar prefers the /u/ of /tut/ to be produced with a particular F_2 value at the center of the vowel, this vowel will always be produced with such a value. This is of course not how phonetic production works – we expect a range of F_2 values to be produced, and for this range to be constrained by the grammar in some way.

One on hand, it is unfair to criticize Flemming's model in this way because the goal of the model was to show how phonological and phonetic phenomena can be accounted for in the same module and not to account for variable phonetic forms. Furthermore, we will see in the next section how stochastic evaluation of constraint rankings (or weightings) can account for variation, and so Flemming's model could be altered to account for variation. However, I have not seen any use of the model with stochastic evaluation, and hence it has not been demonstrated that the model can indeed account for the types of variation that occur naturally with phonetic productions. Even if the fundamental critique given by Boersma, that an intermediate phonological form is necessary, can be overcome, the model as is would need to be modified so that it can generate variable phonetic outputs. I see no reason to attempt such modifications when there is a proposed model that can account for both variation and the influence of phonetics on phonology. This model is discussed in detail in the next section. The possibility of using Flemming-style constraints in a multi-level model is discussed in §6.1.4.

4.3 Model of Analysis: Boersma's Bidirectional Stochastic OT

Boersma's (1997, 1998, 2006, 2007a,b) model, Bidirectional Stochastic OT,[40] is unlike Keating's model in that it does not assume a complete separation of phonetics and phonology but like Keating's model in that it makes use of both discrete and continuous output forms. I believe that, in many ways, Boersma's model retains the spirit of the Window Model, but, as we will see, it is advantageous in that it can not only account for

[40] There are variations of the model in the different Boersma papers, and so, where the papers differ, I follow the model as it is presented in Boersma (2007a).

variation but can control that variation in a way the window model cannot. In order to understand the workings of this particular model and how it differs from classic OT, there are several areas that need to be addressed. This section is organized as follows. I first discuss stochastic evaluation in §4.3.1 and then present the Bidirectional Stochastic OT model in §4.3.2. We focus on the part of the model that accounts for the phonetics-phonology interface in §4.3.3. The Gradual Learning Algorithm (GLA), which is used to develop Stochastic OT rankings, is presented in §4.3.4, and the use of PRAAT in coordination with the GLA and Stochastic OT is discussed in §4.3.5. A brief review of the use of Bidirectional Stochastic OT in the literature is presented in §4.3.6.

4.3.1 Stochastic Evaluation

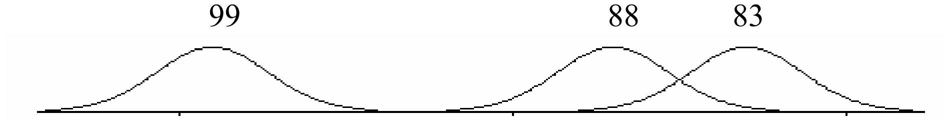
In OT, EVAL is the function that selects an optimal candidate on the basis of a constraint ranking. Each time this function is performed is a *point of evaluation*. A constraint ranking can be defined by the *ranking value* of each constraint. If C_1 has a larger ranking value than C_2, then C_1 » C_2. In classic OT, constraints have the same ranking value at every point of evaluation, as shown in (4.3). Because dominance relations are derived by simply ordering the constraints in accord with their ranking values, the distance between ranking values is not meaningful. For these reason, OT analyses usually reference dominance relations instead of ranking values.

(4.3) C_1 » C_2 » C_3

C_1	»	C_2	»	C_3
99		88		83

In Stochastic OT (StOT), a constraint's rank is defined by a *mean ranking value*. At each point of evaluation, statistical noise is added to each constraint's mean ranking value to get the ranking value used for the constraint at that point of evaluation. I will call this value (meaning ranking value + noise) an *evaluation ranking value*. It is important that the mean ranking value does not change in the process of evaluation. The evaluation ranking value is used for one point of evaluation only; at the next point of evaluation, noise is again added to the mean ranking value to get a new evaluation ranking value. Because the statistical noise added to a given mean ranking value is drawn from a normal distribution, the evaluation ranking values of a given constraint follow a normal distribution, as shown in (4.4).[41] In this hypothetical grammar, the mean ranking value of C_1 is considerably larger than the mean ranking value of C_2, and so there is almost zero probability that C_2 will dominate C_1 at any point of evaluation (i.e. there is almost zero probability that C_2 will have a higher evaluation ranking value than C_1). C_2 and C_3, on the other hand, have mean ranking values that are close enough that, at some points of evaluation, C_2 will dominate C_3, and, at other points of evaluation, C_3 will dominate C_2. In both classic OT and StOT, the distance between the evaluation ranking values of two constraints does not in any way affect the dominance relation. If the constraints in (4.4) have the evaluation ranking values $C_1 = 99.7$, $C_2 = 84.6$, $C_3 = 81.9$, then $C_1 \gg C_2 \gg C_3$ at that point of evaluation, and the fact that the difference between C_1 and C_2 is greater than that between C_2 and C_3 does not influence the rest of EVAL. However, unlike in classic OT, in StOT, the distance between *mean* ranking values of constraints is important because this distance defines the likelihood of one constraint dominating another.

(4.4) C_1 » C_2 »/« C_3
99 88 83



[41] The value added as noise is pulled from a normal distribution with mean 0. The standard deviation doesn't matter as long as it is always the same. Boersma tends to use a standard deviation of 2 and this is the default of PRAAT, and so this is the value used in the analyses of Chapter 5.

At any point of evaluation in StOT, the constraint ranking, and hence the winning candidate, may differ from another point of evaluation. In this way, StOT can account for intra-language variation. Furthermore, how much variation occurs is defined by the mean ranking values of the relevant constraints. Once we have determined the constraints' mean ranking values, we can determine how often a particular output will be the winner for a particular input by running successive points of evaluation and calculating how often each candidate wins. In this way, a StOT grammar can be used to define the probability that an input will be mapped onto some output, and hence to define a set of probabilistic output distributions for any input.

4.3.2 The Bidirectional Model of Production and Perception

As shown in Fig. 4.1, the Bidirectional Model accounts for both production and perception and makes use of four different representations.[42] An underlying form is a stored abstract lexical entry, and is denoted by vertical lines. A surface form, denoted by slashes, is the output of the phonology, and as such is also an abstract discrete form. The model makes use of two different continuous phonetic forms – the auditory form, which contains acoustic information such as formant and fundamental frequency values, and the articulatory form, which contains information about what the articulators are doing. As discussed above, Boersma motivates the use of all these forms in the model because of their use by native speakers for certain tasks. The Bidirectional Model makes use of ranked (and not weighted) constraints, but the ranking is variable due to stochastic evaluation. Importantly, all

[42] In Boersma (2007b), morphemes are also included at the top of the model and are related to the underlying form via lexical constraints. Thus, a morpheme (or combination of morphemes) is really the input to production and the final output of perception. Because this level is often omitted in Boersma's other work and is not relevant for the analysis presented here, I will ignore it in the rest of this paper.

constraints are bidirectional in that they are applicable in both the task of production and perception (except for the articulatory constraints which only do work in production).[43]

Figure 4.1: Boersma's (2007a: 2031) "bidirectional model of phonology and phonetics"

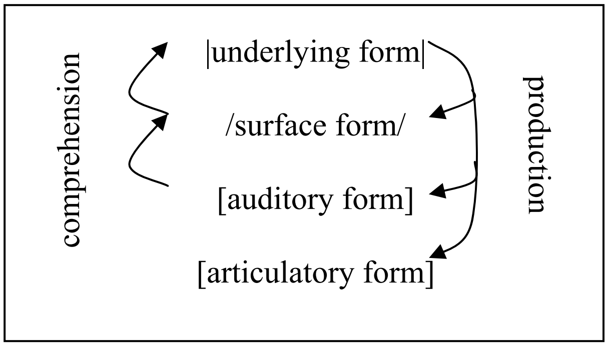
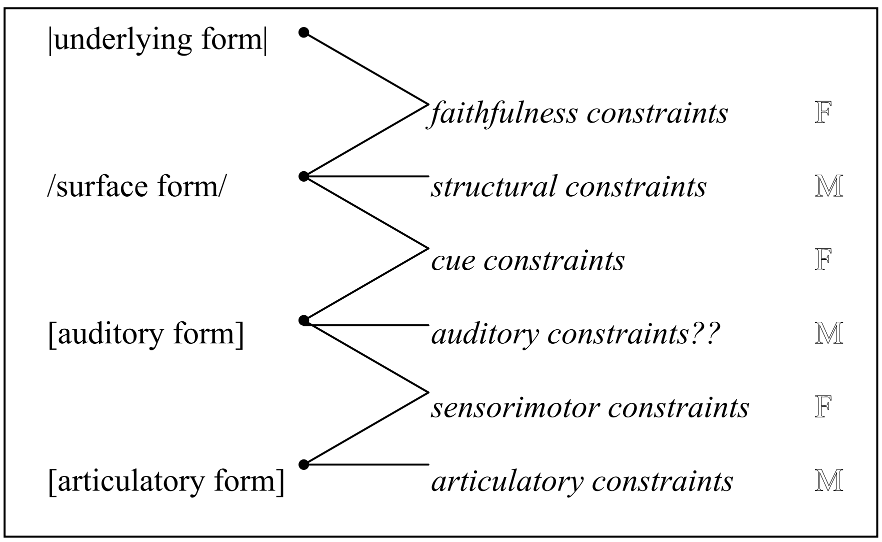


Figure 4.2: Types of constraints used in the bidirectional model (Boersma 2007a: 2032)



The constraints used in the bidirectional model are shown in Fig. 4.2. These constraints are used for both production and perception as applicable. Each constraint acts like either a classic OT faithfulness constraint or a classic OT markedness constraint, as marked by 𝔽 and 𝕄, respectively. Faithfulness constraints compare two forms (either an input and output form or two output forms) and assign violation marks accordingly, while markedness constraints look just at an output candidate and assign violation marks

[43] See Smolensky (1996) for an earlier use bidirectional faithfulness constraints and Tesar (1997) and Tesar and Smolensky (1998, 2000) for earlier uses of bidirectional structural constraints.

accordingly.

In the production grammar, an underlying form is the input, and candidate triplets composed of a surface, auditory, and articulatory form compete as candidates, as illustrated in Fig. 4.1. *Structural constraints* and *articulatory constraints* look at just the surface and articulatory forms (respectively) and assign violation marks for any offending characteristics of those forms. Boersma (2007a) hypothesizes that there may be *auditory constraints* that penalize certain properties of auditory forms (such as constraints against loud or irritating noises) but does not make use of such constraints. The possible existence of auditory constraints will not be discussed further in this dissertation. All markedness constraints ignore the input and the other irrelevant forms in the candidate triplet. *Faithfulness constraints* compare the input to the surface form in each candidate triplet and assign violation marks as applicable. *Cue constraints* compare two forms in each candidate triplet: the surface and auditory forms, while *sensorimotor constraints* compare two different forms of the candidate triplet: the auditory and articulatory forms. Each of these faithfulness-type constraints assigns violation marks for relevant mismatches between the two forms of comparison. In this way, all of the constraints from Fig. 4.2 are used in production in one pass through the system, which makes use of both categorical and gradient outputs.

In the perception grammar, there are two stages, as illustrated in Fig. 4.1. The same constraints with the same mean ranking value are used in both production and perception. In the first stage, called *perception* or *prelexical perception* (Boersma 2006: 169), the listener hears an auditory form, which is thus the input to the grammar. Surface forms compete as output candidates. Structural constraints assign violation marks to offending surface forms, while cue constraints assign violation marks to surface forms that should not be paired with the input acoustic form as dictated by the constraint. In the second stage, called *word recognition* (Boersma 2006: 170), the surface form that was the winner from the first stage is the input. Underlying forms compete as output candidates. The winner is determined by the ranking of the faithfulness constraints, which compare the input and output and assign violation marks accordingly.

For practical purposes, the model in Fig. 4.1 can be simplified somewhat because the relationship between auditory and articulatory forms is proposed to be universal. Boersma (2007a: 2032) says "The relation between ArtF and AudF … is the usual universal sensorimotor mapping and could be described with constraints as well…"
Thus, while we could define the relation with constraints, this relation is not language-specific and so it is not crucial for developing a language-specific grammar. If we ignore the sensorimotor constraints and conflate the two continuous forms in one [phonetic form], we get the simplified model in Fig. 4.3 and the constraints in Fig.4.4. In this simplified model, the phonetic form acts as the auditory form in that it is the input to perception and its relation to the surface form is regulated by cue constraints. Though articulatory constraints are supposed to penalize *articulations*, in this model we can think of the articulatory constraints as penalizing auditory properties that are caused by certain articulations. For example, in the full model, the articulatory constraint *[vibrate vocal folds] would penalize an articulatory form with vocal fold vibration. In the simplified model, this constraint penalizes a phonetic form with the auditory property of voicing because this is the auditory form that is universally linked with an articulatory form with vibrating vocal folds.

Figure 4.3: A simplified bidirectional model

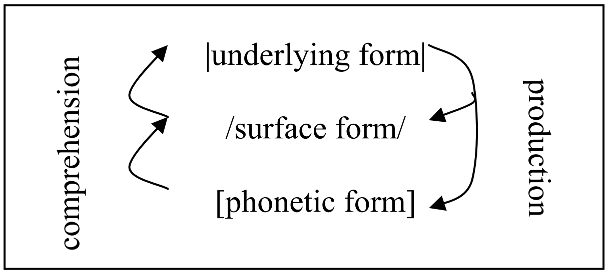
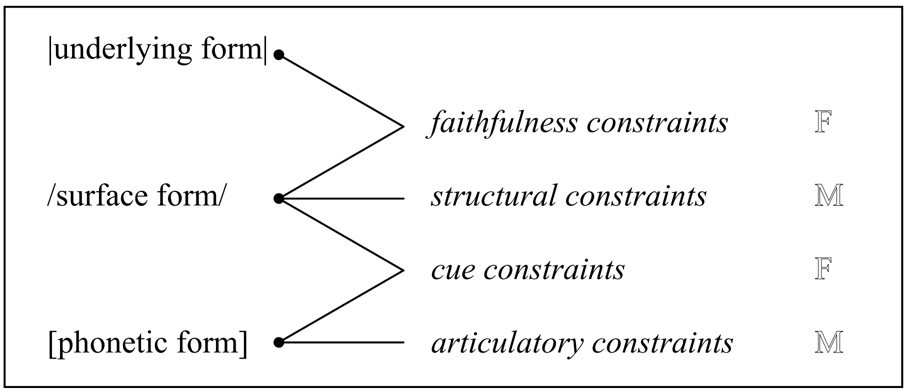


Figure 4.4: Types of constraints used for the simplified model



The basic properties of this simplified model are illustrated with a hypothetical grammar that is defined by the specific constraints in (4.5). The tableau in (4.6) makes use of these constraints as ranked for some point of evaluation. In this tableau, the winning candidate has a surface form that is faithful to the underlying form (as demanded by the high-ranking ID[low]) and a phonetic form with an F_1 value that incurs a violation from the lowest-ranked cue constraints. In this way, we see how the different forms in a candidate set are regulated by different constraints and also how these different constraints interact with each other. For example, if *FULLV dominated ID[low], then not only would a different surface form (/ə/) win, but also a different phonetic form ([F_1 = 500 Hz]) because this different phonetic form is the preferred pairing with the surface form /ə/.

(4.5) constraints used in a hypothetical grammar[44]

faithfulness constraint:

ID[low]

an underlying form of |a| should be paired with a surface form of /a/

structural constraint:

*FULLV

no full vowel should appear in a surface form

articulatory constraint:

*[F_1=700 Hz]

an F_1 of 700 Hz should not appear in a phonetic form

cue constraints:

*/a/, [F_1 = 500 Hz]

a surface form of /a/ should not be paired with a phonetic form with [F_1 = 500 Hz]

*/a/, [F_1 = 600 Hz]

a surface form of /a/ should not be paired with a phonetic form with [F_1 = 600 Hz]

*/a/, [F_1 = 700 Hz]

a surface form of /a/ should not be paired with a phonetic form with [F_1 = 700 Hz]

*/ə/, [F_1 = 500 Hz]

a surface form of /ə/ should not be paired with a phonetic form with [F_1 = 500 Hz]

*/ə/, [F_1 = 600 Hz]

a surface form of /ə/ should not be paired with a phonetic form with [F_1 = 600 Hz]

*/ə/, [F_1 = 700 Hz]

a surface form of /ə/ should not be paired with a phonetic form with [F_1 = 700 Hz]

(4.6) example production tableau for phonology and phonetics

|a|	ID[low]	*/a/,[F_1 = 500 Hz]	*/a/,[F_1 = 700 Hz]	*[F_1 = 700 Hz]	*/ə/, [F_1 = 700 Hz]	*/ə/, [F_1 = 600 Hz]	*FULLV	*/ə/, [F_1 = 500 Hz]	*/a/,[F_1 = 600 Hz]
/a/ [F_1=500 Hz]		*!					*		
☞ /a/ [F_1=600 Hz]							*		*
/a/ [F_1=700 Hz]			*!	*			*		
/ə/ [F_1=500 Hz]	*!							*	
/ə/ [F_1=600 Hz]	*!					*			
/ə/ [F_1=700 Hz]	*!			*	*				

[44] The articulatory and cue constraints reference F_1 values, but we know that F_1 values for the same vowel differ by speaker. Such constraints are used for simplicity in our hypothetical example, but in reality these constraints would need to reference relativized F_1 values to account for speaker normalization.

Even though the model is not modular, it can be used for just phonology or just phonetics (see more on the latter in §4.3.3). If the phonologist is concerned about what phonological output |a| is mapped onto, this can generally be done by ignoring the phonetic forms and all constraints but the faithfulness and structural constraints.[45] If the phonetician is concerned about what acoustic values for F_1 are produced for /a/, this can be done by ignoring the underlying form and the faithfulness and structural constraints. Importantly, though, because the actual candidates in production tableaus are sets of discrete and continuous forms, the model allows for the influence of phonetics on phonology, and hence can account for situations where phonetic concerns affect phonological decisions (as explained in detail in Boersma 2007a).

The two stages of perception are illustrated with our hypothetical grammar in (4.7) and (4.8). In the first stage in (4.7), the input of [F_1 = 600 Hz] is mapped unto the surface form /a/. Not all constraints that were used in production are relevant at each stage of perception. In the following tableaus, I show all the constraints of the grammar, and the shaded constraints are those that are not able to assign any violation marks in a particular tableau. In this case, the winning candidate is determined solely by the ranking of the structural constraint and the relevant cue constraints. Because it is more costly to perceive an F_1 of 600 Hz as /ə/ than it is to perceive a full vowel or the same F_1 value as /a/, /a/ is the winner.

[45] Boersma (2007b: 2) is clear on this point, saying "For most phonologists, the surface form does not contain any concrete continuous phonetic detail, and this is something I agree with; it means that one can do insightful investigations in many areas of phonology by just considering the two discrete phonological representations."

(4.7) example tableau for prelexical perception

[F_1=600 Hz]	ID[low]	*/a/,[F_1 = 500 Hz]	*/a/,[F_1 = 700 Hz]	*[F_1 = 700 Hz]	*/ə/, [F_1 = 700 Hz]	*/ə/, [F_1 = 600 Hz]	*FULLV	*/ə/, [F_1 = 500 Hz]	*/a/,[F_1 = 600 Hz]
/ə/						*!			
☞ /a/							*		*

In the second stage, the winner /a/ is now the input. Only faithfulness constraints are relevant at the stage of word recognition. Because our hypothetical grammar only contains the faithfulness constraint ID[a], the input of /a/ is always mapped onto the underlying form |a|.

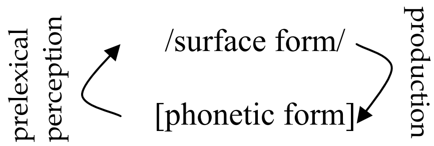
(4.8) example tableau for word recognition

/a/	ID[low]	*/a/,[F_1 = 500 Hz]	*/a/,[F_1 = 700 Hz]	*[F_1 = 700 Hz]	*/ə/, [F_1 = 700 Hz]	*/ə/, [F_1 = 600 Hz]	*FULLV	*/ə/, [F_1 = 500 Hz]	*/a/,[F_1 = 600 Hz]
|ə|	*!								
☞ |a|									

4.3.3 Specific Model of Analysis Used in this Paper: The Bidirectional Model at the Phonetics-Phonology Interface

We will now look at how the phonetics-phonology interface is modeled by the relationship between the surface form and the acoustic/articulatory forms. Again, for simplicity, I conflate the acoustic and articulatory forms into one [phonetic form]. This means that the [phonetic form] is both the input to first stage of the perception grammar and under the jurisdiction of the articulatory constraints in the production grammar. The simplified model (which is similar to the model used in Boersma 2006) is illustrated in (4.9) and shows how the production grammar uses a discrete surface form to generate a continuous phonetic form that the speaker actually says, while the perception grammar uses the continuous phonetic form that is heard to identify a discrete surface form.

(4.9) simplified Bidirectional model (phonetics-phonology interface only)



Cue constraints, such as those defined in (4.5), are relevant in both production and perception since they penalize the pairing of surface forms and phonetic forms without regard for which form is the input and which form is the output candidate. Structural constraints are only relevant in perception, where surface forms are the output candidates. Finally, articulatory constraints are relevant to the production grammar, as they assign violation marks to offending phonetic forms.

In order to illustrate how these constraints are used in the production and perception grammar, example tableaus are given below. In (4.10) we see an example tableau for production. Here the input /a/ is mapped onto the winning candidate with an F_1 of 600 Hz. The losing candidates crucially violate cue constraints that penalize the pairing of /a/ with their specific F_1 values. The last candidate also violates the articulatory constraint *[F_1 = 700 Hz]. Note that no candidate can violate the structural constraint *FULLV because this constraint penalizes surface forms which are not the candidates in a production tableau.

(4.10) example production tableau

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

The perception tableau in (4.11) shows how the cue constraints are used in the perception grammar. Additionally, structural constraints assign violation marks to offending surface forms. In this grammar it is the articulatory constraints that have no effect because they assign violation marks to phonetic forms, which are not the candidates in a perception tableau.

(4.11) example perception tableau

$[F_1=600\text{ Hz}]$	*/u/, $[F_1=600\text{ Hz}]$	*/o/, $[F_1=600\text{ Hz}]$	*$[F_1=600\text{ Hz}]$	*FULLV	*/a/, $[F_1= 600\text{ Hz}]$
☞ /a/				*	*
/o/		*!		*	
/u/	*!			*	

The above tableaus illustrate how the Bidirectional StOT model can be simplified for use with the phonetics-phonology interface. This model should be usable by any OT researcher as it makes use of constraints that act like either faithfulness or markedness constraints and output candidates with only one form. The primary differences from classic OT are that stochastic evaluation is a necessity (in order to account for variation) and that phonetic forms are continuous. The practical consequence of this last fact is that many more constraints are needed to do insightful analysis. The implications of this are discussed in §§5.2 and 6.1.1.

4.3.4 The Gradual Learning Algorithm

We have now seen how the production and perception grammars work in the Bidirectional Model. Before using this model for analysis, we have yet to determine how to develop a StOT constraint ranking. This is a nontrivial issue as StOT rankings are more complex than classic OT rankings. It is not enough to simply derive a dominance relation; we must identify mean ranking values that can be used to define the probability of possible dominance relations. For this reason, it is useful to employ an algorithm for determining the mean ranking values of each constraint that most accurately account for the data.

The Gradual Learning Algorithm (GLA, Boersma and Hayes 2001, Boersma 1997) can be used to develop a StOT ranking. This algorithm is a model of language acquisition such that it models how the learner adjusts an interim constraint ranking when faced with data that contradicts that ranking. The GLA is illustrated in (4.12) and details follow.

(4.12) adjustment of mean ranking values with the GLA (Boersma and Hayes 2001)
✋ denotes the learning datum; ☞ denotes the winning candidate

	→	→		←	→	←
/surface form/	C_1	C_2	C_3	C_4	C_5	C_6
✋ [phonetic $form_1$]	*	*			*	
☞ [phonetic $form_2$]				*		*

The scenario illustrated in (4.12) is one in which the learner just heard someone say [phonetic $form_1$] for /surface form/. Thus, [phonetic $form_1$] is the learning datum, marked with (✋). However, the current interim grammar predicts [phonetic $form_2$] to win, given the input of /surface form/. The winning candidate is denoted by the traditional pointing finger (☞). In this case the learner's current grammar predicts an incorrect winner. In order to modify the grammar in favor of the learning datum, all constraints that favor the incorrect winner over the learning datum are demoted (such that their mean ranking values are decreased by a small amount, called the *plasticity*) and all constraints that favor the learning datum over the incorrect winner are promoted (such that their mean ranking values are increased by the plasticity). As this process continues, mean ranking values of constraints are promoted and demoted in order to better fit the data.

When using the GLA, there are some variables that each linguist must define: the initial state of the grammar, the plasticity of constraint adjustment, and when to stop adjusting constraints.

The *initial state* of the grammar is defined by the mean ranking values of each constraint before any adjustments are made by the GLA. It is thus up to the linguist to use whatever initial state is preferable on the basis of hypotheses about language acquisition. Interestingly, Boersma and Hayes (2001) point out that, in the set of analyses discussed in that paper, the initial state of the grammar had no effect on the final grammar developed through the GLA. Because the initial state may very well be irrelevant for the GLA, it is common to use an initial state where all constraints have the same ranking value.

The *plasticity* defines what value is added to or deducted from a constraint's mean ranking value in the event that the current grammar predicts an incorrect winner. In theory, the plasticity can be any amount the linguist wants. However, Boersma and Hayes explain that the plasticity should be rather small:

> "A small plasticity value does a better job of matching learning data frequencies in the end, but a large plasticity value nears its goal faster. The virtues of these two approaches can be combined by adopting a learning schedule that decreases the plasticity as learning proceeds. This seems in principle realistic: in humans, grammar apparently stabilizes in adulthood, as nonlexical learning slows or halts." (Boersma and Hayes 2001: 79)

Finally, the algorithm must stop at some point. The linguist can either terminate the algorithm after a set number of trials or after the grammar is correct a set percentage of the time.

All of these variables that affect the procedure of the GLA are discussed in more detail in the next section when their usage by PRAAT is presented.

It is clear that the GLA is a significant advancement not just for phonetics and phonology but also for theories of language acquisition. For this reason, a general disclaimer is relevant at this time. For the purposes of this paper, I use the GLA primarily as a tool for developing StOT rankings (which would not be possible by hand). I do not take any strong stance as to how accurately the GLA reflects real scenarios of language acquisition. It should also be pointed out for any researchers who are more interested in the computational side of the issue that a weakness of the GLA has been exposed by Pater (2008).[46] However the particular problem identified by Pater has not come up in the analyses presented in Chapter 5. Furthermore, Magri (2008) shows how to modify the GLA so that Pater's problem is solved. Though the software I use to run the GLA does not implement Magri's solution, the fact that Pater's problem does not relate to my data leaves me confident that the GLA is appropriate for my uses. See footnote 10 in Boersma and Hamann (2008) for further discussion of Pater's critique.

4.3.5 Stochastic OT, the GLA, and PRAAT

The previous sections have explained how the Bidirectional StOT model works and how the GLA is used to develop StOT rankings. For those who wish to do research in this area, PRAAT contains all the necessary programming to develop StOT rankings through the GLA. In this section I will briefly explain how the linguist can use PRAAT to create StOT grammars (see the PRAAT manual (or Boersma 1999a for a printer-friendly version) for more details on this topic). I will explain the types of information that are given to PRAAT through the use of another hypothetical example. For ease of exposition, we will do a phonological analysis (of the mapping of an underlying representation onto a surface form) and ignore phonetic forms.

Consider two different languages: one where |vn| (a vowel – nasal stop sequence) categorically maps onto /ṽn/, and another where |vn| often maps onto /ṽn/. In both

[46] See Pater (2008) for the specifics of his critique of the GLA. Briefly, the problematic situation is shown in the comparative tableau below (identical to Pater 2008: (6)). In this case the only ranking compatible with the data is $C_1 \gg C_2 \gg C_3 \gg C_4 \gg C_5$, but the GLA does not converge on this grammar, and instead continually pushes the ranking values of each constraint higher and higher.

		C_1	C_2	C_3	C_4	C_5
In_1	Out_1-W ~ Out_1-L	W	L	W		
In_2	Out_2-W ~ Out_2-L		W	L	W	
In_3	Out_3-W ~ Out_3-L			W	L	W
In_4	Out_4-W ~ Out_4-L				W	L

languages, |v| maps onto /v/. We will analyze this language with the constraints in (4.13).

(4.13)	ID[nasal]	*input and output forms should be identical for the feature [nasal]*
	*/vn/	*an oral vowel should not be followed by a nasal consonant in a surface form*
	*/ṽ/	*vowels should not be nasalized*

We need to let PRAAT know what inputs to evaluate, what candidates to generate for each input, what constraints to use, how those constraints assign violation marks, and what the desired pairings of inputs and outputs are. This is done through use of two different files: a "pair distribution file" and an "OT grammar file". The first file specifies how often a particular candidate is the winner (in the target grammar) for a particular input, as shown in (4.14). If one candidate always wins and the other always loses, we can represent this with 1's and 0's. If we wanted to denote a specific level of variation, say |vn| → /ṽn/ 75% of the time and |vn| → /vn/ 25%, we would just adjust the proportions accordingly, i.e. with |vn| → /ṽn/ = 75 and |vn| → /vn/ = 25.

(4.14) pair distributions

categorical assimilation		variable assimilation	
|vn| → /ṽn/	= 1	|vn| → /ṽn/	= 75
|vn| → /vn/	= 0	|vn| → /vn/	= 25
|v| → /ṽ/	= 0	|v| → /ṽ/	= 0
|v| → /v/	= 1	|v| → /v/	= 1

In the OT grammar file, we specify the constraints to be used, their initial ranking, and the tableaus to be evaluated. In (4.15) we see that each of the constraints is set with an initial ranking of 100 and that there are two tableaus to be evaluated. For each candidate in a tableau, we see how many violation marks are assigned by C_1-C_3. With this information, PRAAT can proceed to adjusting the ranking values in accord with the GLA so that the resulting grammar predicts the correct distribution of |vn| → /ṽn/ mappings.

(4.15) initial state of grammar

C_1: ID[nasal]	100		
C_2: */vn/	100		
C_3: */ṽ/	100		
tableau$_1$:			
input: |vn|			
cand$_1$: /vn/	0	1	0
cand$_2$: /ṽn/	1	0	1
tableau$_2$:			
input: |v|			
cand$_1$ /v/	0	0	0
cand$_2$: /ṽ/	0	0	1

I used the grammar in (4.15) with each of the pair distributions in (4.14) to run learning simulations with PRAAT, using the default settings for plasticity and other variables. For the "categorical" distribution, we get a mean ranking value of 106.9 for */vn/ and 93.1 for ID[nasal] and */ṽ/. With these mean ranking values, |vn| should be mapped onto /ṽn/ almost 100% of the time, as shown in (4.16a). I used PRAAT to run 100,000,000 points of evaluation, and |vn| was mapped onto /vn/ 112 times, for a rate of .000112%. In (4.16b), /ṽ/ is harmonically bounded and will never win no matter what the constraint ranking is.

(4.16) a. |vn| → /ṽn/ almost 100% of the time

|vn|	*/vn/	*/ṽ/	ID[nasal]
☞ /ṽn/		*	*
/vn/	*!		

b. |v| → /v/ 100% of the time (/ṽ/ is harmonically bounded)

|v|	*/vn/	*/ṽ/	ID[nasal]
/ṽ/		* (!)	* (!)
☞ /v/			

With the variable distribution the mean ranking values as learned with the GLA were closer together: 101.4 for */vn/ and 98.6 for ID[nasal] and */ṽ/. With these mean ranking values, there is a 25% chance that either ID[nasal] or */ṽ/ will dominate */vn/ and hence that /vn/ will be the optimal output, as shown in (4.17). Out of 100,000 simulations of this tableau, /ṽn/ won 75, 274 times, and /vn/ won 24,726 times.

(4.17) |vn| → /ṽn/ 75% of the time; |vn| → /vn/ 25% of the time

|vn|	*/vn/	*/ṽ/	ID[nasal]
75% ☞ /ṽn/		* (!)	* (!)
25% ☞ /vn/	*(!)		

The default settings of PRAAT use 4 levels of plasticity (1.0, then 0.9, then 0.8, then 0.7) with 100,000 replications per plasticity. This means that after 400,000 replications the program is terminated. The user may change any of these settings. In this paper, I use the default settings with regard to plasticity and program termination unless otherwise stated.

4.3.6 Applications of the Bidirectional StOT Model

The literature on Bidirectional StOT has focused mainly on the relation between cues that are controlled by the speaker in production and cues that are used by the listener in perception. The model has also been used to account for the "prototype effect" and its interaction with dispersion theory. In this section I will briefly summarize some of the literature that has applied Bidirectional StOT to account for linguistic phenomena.

As summarized in Chapter 1, there is plenty of evidence of symmetries in production and perception, and this evidence supports the intuition behind cue constraints. If the goal of the speaker is to be understood, then the cues that are most likely to lead to correct perception should be the same cues that are emphasized in production. Similarly, if the goal of the listener is to understand, then the listener should attend to those cues that are emphasized in production. The results of production and perception studies indicate that speakers and listeners do behave in this way and a main advantage of the Bidirectional Model is that cue constraints elegantly describe this behavior.

It was also mentioned in Chapter 1 that Escudero and Boersma (2004) have used cue constraints to analyze the use of the cues of F_1 and length in the tense/lax contrast in different English dialects (see also Escudero 2001, 2002). Additionally, the ways that cue constraints interact with other components of the model are described in detail in Boersma (2007b). This paper includes several language specific analyses (Japanese, Korean, English, among others) to show how the various relations between production and perception in these languages can be accounted for with the model.

Articulatory constraints penalize effortful productions, and so they are used to penalize the production of acoustic values that are good cues in perception. In this way, articulatory constraints can be used to account for mismatches between production and perception. One such mismatch is known as the "prototype effect", which is analyzed by Boersma (2006). In English, listeners prefer an /i/ with a very high F_2 and a very low F_1, but the majority of /i/ tokens in production have less peripheral formants. There is thus a difference between the most commonly produced tokens and the best perceptual tokens. (See Boersma 2006 for a full review of the literature that documents and analyzes this phenomenon.)

Articulatory constraints can be used to explain why the majority of produced /i/'s have less space between F_1 and F_2. If very low F_1's and very high F_2's are costly productions, then the relevant articulatory constraints can assign fatal violation marks to candidates with a very low F_1 or a very high F_2, even if these phonetic values are preferred by the ranking of cue constraints. It is thus clear why the production data has the distribution it does, but why do listeners prefer tokens that are only marginally present in the production data? Boersma explains that, under the terms of his model, the protypicality judgment task involves only the mapping of the surface form onto the auditory form (see Fig. 4.1). The participant is told something like, "You are going to hear an /i/; tell me how good it is." So, the participant has some surface form in mind and is asked to judge the auditory form on the basis of that surface form. In order to perform this task, the participant need only map the surface form onto the auditory form. Because the participant is not actually performing the task of production, the auditory form need not be mapped onto an articulatory form, and so the articulatory constraints never come into play. The participant is thus going to select the auditory form with the best cues regardless of how effortful these cues are. In a broader sense, this analysis captures the intuition that the listener's ideal situation is one in which the speaker makes every possible effort to produce a clear utterance because this is less effortful for the listener.

Boersma and Hamann (2008) take this type of analysis a step further by showing how the "prototype effect" competes with an "articulatory effect" to derive generalizations about auditory dispersion. It has been demonstrated that the sounds in the phoneme inventory of a language tend to be optimally auditorily distinct.[47] Boersma and Hamann use the Bidirectional Model to account for the diachronic development of such sound systems. They say (regarding sibilants), "While the bidirectional use of cue constraints causes the categories to drift apart auditorily, the presence of the articulatory constraints checks this expansion and drives the production distributions back towards the centre of the spectral mean continuum (p. 244)."

The Bidirectional Model has thus been used successful to account for a variety of patterns that occur in natural language. There is one thing missing from the literature, though. In all of the papers summarized above, when the analysis is of the phonetics-phonology interface, the data of analysis is not a set of tokens of natural speech, but rather a set of idealized surface form – phonetic form pairings based on known phenomena. Thus, using the example of the tense/lax distinction in English, we know the patterns that occur in production, and these patterns were used to define a set of data that is analyzed with the Bidirectional Model. But in a sense this data is too perfect; the real child language learner must develop a grammar from data that is inconsistent and imprecise. The question is then whether or not the Bidirectional Model works if constraint rankings are developed to match real (but imperfect) data. That is exactly the goal of this paper, to use a set of actual language productions to develop a Bidirectional Stochastic OT grammar. This paper will show that the model can indeed account for the variable and less than ideal data that a real language learner would be exposed to, which motivates future research in this area.

[47] Boersma and Hamann summarize a body of literature on how the components of phoneme inventories are maximally auditorily distinct. For example, languages that have three high vowels tend to have one with a higher F_2 (e.g. [i]), one with a mid F_2 (e.g. [ɨ]), and one with a lower F_2 (e.g. [u]), such as in Polish (Jassem 2003). Languages with only one high vowel tend to have a high vowel with a mid F_2, such as Kabardian with only [ɨ] (Choi 1991).

4.4 Chapter Summary

In this chapter we have looked at various proposals for the analysis of the phonetics-phonology interface. The classic view, as exemplified by Keating's Window Model, is that phonetics follows phonology serially. A different view of the phonetics-phonology interface is that all phonological and phonetic processes are done simultaneously. This is the claim of two-level models, like Flemming's Unified Model. In this model, a categorical underlying form is mapped onto only a discrete phonetic form. However, Boersma (2007a) points to theoretical flaws with both of these models as the serial model does not allow for phonetics to affect phonology and the two-level model ignores an intermediate phonological form that is necessitated by its use in perceptual tasks.

A third approach to the phonetic-phonology interface is the one that will be used in this paper: Boersma's Bidirectional Stochastic OT. This model uses multiple levels of representation, but all processes are done in a single module. Importantly, the model makes use of bidirectional constraints that have the same mean ranking values in production and perception. Stochastic evaluation of constraint rankings can account for variation in the optimal outputs associated with a given input. The GLA can be used to develop StOT rankings. In the next chapter, I use Bidirectional StOT at the phonetics-phonology interface to define a grammar that accounts for the relationships between the cues of pitch and glottalization to the HIGH TONE and GLOTTALIZED vowels of Yucatec Maya. Data from the production studies (chapter 2) is used to simulate the data pool that the learner of Yucatec Maya would be exposed to and the GLA is used to learn a constraint ranking that accounts for the data.

5

THE PRODUCTION AND PERCEPTION GRAMMAR OF YUCATEC MAYA

In this chapter we will use data from the production experiments to mimic the data that the child language learner would be exposed to in order to develop a grammar through the GLA (see §4.3.4). Because we are using bidirectional constraints, this grammar is both a production and a perception grammar. Our target grammar is one that accounts for the production and perception of pitch and glottalization with HIGH TONE and GLOTTALIZED vowels, and so the other two vowel shapes are excluded from analysis. Furthermore, as discussed in Chapters 2 and 3, we wish to focus on a single dialect group so that dialect variation will not influence the grammar. The dialect of consideration for analysis is that spoken in Santa Elena, Yucatan, Mexico (see map in Fig. 1.1). As will be discussed in §5.1, not all of the data from Santa Elena that was presented in chapter 2 will be included in the data set that is used as input for the GLA. Because the GLA is a model of language acquisition, I want to focus on only that data that would constitute the data pool that a single child might be exposed to. It is for this reason that other dialect groups are excluded, and other data is excluded on this basis as discussed in §5.1.

In addition to the goal of modeling the production and perception of pitch and glottalization in Yucatec Maya, another goal of this chapter is to assess the success of Boersma's Bidirectional Model (see §§4.3.2 and 4.3.3). We will see that the framework of the model, including the specific types of cue and articulatory constraints proposed by

Boersma, can indeed account for the Yucatec Maya data. We will, however, have to adjust the exact learning strategy proposed by Boersma (2006, see §5.3). While Boersma proposes that the rankings of cue constraints are learned through training on only perception tableaus, we will see that production tableaus must also be used in order for the learner to reach an accurate adult grammar. Furthermore, some language-specific learning is also necessary for the articulatory constraints. These results lead to the conclusion that the Bidirectional Model can accurately account for real language data but only when the learner is trained on both production and perception tableaus.

This chapter is organized as follows. I first review the relevant data that we want the grammar to account for in §5.1. This review will focus on determining just which phonetic dimensions can be used to define the contrast between HIGH TONE and GLOTTALIZED vowels, as these are the likely dimensions that are controlled by the speaker and/or attended to by the listener. Based on this discussion of the data, I propose cue and articulatory constraints that are relevant for our model in §5.2. The learning strategy proposed by Boersma (*lexicon-driven perceptual learning*) as well as alternative learning strategies are discussed in §5.3. In this section we look at hypothetical data that is problematic for learning through perception tableaus alone, and I propose that production learning may be necessary to account for such data.

In the rest of the chapter different learning strategies are used to develop adult grammars of Yucatec Maya. I use perception learning to develop ranking values of cue constraints in §5.4. We will see that this learning strategy leads to a highly accurate perception grammar but a faulty production grammar. Articulatory constraints are then added to the model in §5.5, and their ranking values are learned through production learning.

Even with articulatory constraints, there are still some problematic areas of the production grammar, and so I add a learning step that uses production tableaus to test the ranking values of both cue and articulatory constraints in §5.6. The grammar that is developed in this section loses a bit of perception accuracy but gains a great deal of production accuracy.

5.1 Review of Data: Which Phonetic Dimensions Signal Contrast?

In this section we first review the production data that was collected from speakers from Santa Elena. This data will be used to make generalizations about the patterns that occur in production. It will also be used as the data that the GLA learner is trained on, as described in more detail in §§5.3 and 5.4. I also review the results of the perception study in §5.1.2.

5.1.1 Production Data

In chapter 2, we saw a variety of details about the production of vowel shape by speakers from different towns in Yucatan, Mexico. As was mentioned at that time, we will focus on a subset of this data for the purposes of developing a grammar of Yucatec Maya. Only data from speakers from Santa Elena will be used as this is the only dialect group from which I was able to record enough speakers to have a robust set of data and because this dialect produces pitch and glottalization in a way that closely resembles the phonological claims in previous literature on Yucatec Maya (see §1.2.2). Due to dialect variation in the production of pitch and vowel length, I will not use data from multiple towns for fear that this would be an inaccurate representation of the linguistic situation that a language learner would be exposed to.[48] Furthermore, I will only use data from two environments: words

[48] It was a difficult decision to exclude data from speakers from Mérida. The reasons for inclusion would be that there are not that many differences between Santa Elena and Mérida, these are the two "western" towns, and more data is always welcome. I chose to exclude data from Mérida because I want to be absolutely certain

spoken in isolation in production study 1 and words spoken in phrase-final position in production study 2 (frame sentence D: "Tu ya'alaj __." *S/he said* __.). These two contexts are used because they are both "phrase-final" and because the data obtained from these two contexts was highly similar. Recall that production study 1 included both existing forms and nonce forms and that there were some differences in the production of the two types of words. For this reason, I exclude nonce forms from the analysis here. Additionally, because initial pitch is an important feature that distinguishes GLOTTALIZED from HIGH TONE vowels, I exclude all tokens from which no initial pitch value could be obtained.[49] To summarize, the data under analysis in this chapter comes from words spoken by Santa Elena speakers where an initial pitch value was identifiable; all words are either existing forms in production study 1 or words in phrase-final position in production study 2. There are 440 tokens from production study 1 (215 with a GLOTTALIZED vowel and 225 with a HIGH TONE vowel) and 245 tokens from production study 2 (122 with a GLOTTALIZED vowel and 123 with a HIGH TONE vowel) that meet these criteria, for a total of 685 tokens (337 with a GLOTTALIZED vowel and 348 with a HIGH TONE vowel).

In the rest of this subsection we will review the data that is the target of analysis here. Because this subset of data is slightly different than any of the subsets looked at in Chapter 2 (e.g. nonce forms and tokens missing an initial pitch value were never excluded in Chapter 2), the numbers presented here will not be identical to those presented in Chapter 2.

With respect to vowel length, words spoken in isolation are longer than words spoken in the frame sentence, but there are no statistically significant differences between the vowel

---

to define a grammar on the basis of the input that one child might be exposed to, and the reality is that a child in Santa Elena would (mostly) only be exposed to speakers from Santa Elena.

[49] A total of 28 tokens from production study 1 and 5 tokens from production study 2 were excluded on this basis.

lengths of GLOTTALIZED vowels as compared to HIGH TONE vowels. Because vowel length cannot be used to distinguish these two vowel shapes, it will not be addressed in this analysis.

Table 5.1 presents the distribution of the four main glottalization types as produced for HIGH TONE and GLOTTALIZED vowels in production study 1, production study 2, and in both production studies combined. It is clear from this data that glottalization type is an important factor that distinguishes GLOTTALIZED from HIGH TONE vowels. The results from production study 1 and production study 2 are minimally different; the only notable difference is that, with GLOTTALIZED vowels, more modal voice and less weak glottalization occurs in production study 2. Table 5.1 provides evidence that 1) glottalization is a phonetic dimension that signals a contrast between GLOTTALIZED and HIGH TONE vowels and 2) context (isolation vs. phrase-final) minimally affects only the production of modal voice vs. weak glottalization with GLOTTALIZED vowels.

Table 5.1: Review of distribution of glottalization types

		modal	weak glottalization	creaky voice	glottal stop
GLOTTALIZED	study 1 (n = 215)	52%	19%	27%	2%
	study 2 (n = 122)	65%	10%	25%	1%
	total	56%	15%	26%	2%
HIGH TONE	study 1 (n = 225)	97%	2%	1%	0%
	study 2 (n = 123)	98%	2%	0%	0%
	total	97%	1%	1%	0%

When investigating the average pitch contours of the vowel shapes, we need to account for the interaction among pitch, glottalization, and gender (see §2.2). Fig. 5.1 shows the average pitch contours by vowel shape and gender for each production study; Fig. 5.2 shows these contours as averaged over both production studies. All pitch contours can be categorized as "falling" in that they all end with lower pitch than they start. The GLOTTALIZED vowels, as spoken by females, are the only vowels for which the lowest pitch values occur in the middle of the vowel. As was explained in §2.2, the reason for this is that creaky voice (which is commonly produced during the middle portion of the GLOTTALIZED vowel) causes females to produce pitch that is much lower than their baseline, while the pitch produced during creaky voice is slightly over the baseline of males.

Figure 5.1: Review of pitch
males = 'm'; females = 'f'; HIGH TONE = gray; GLOTTALIZED = black

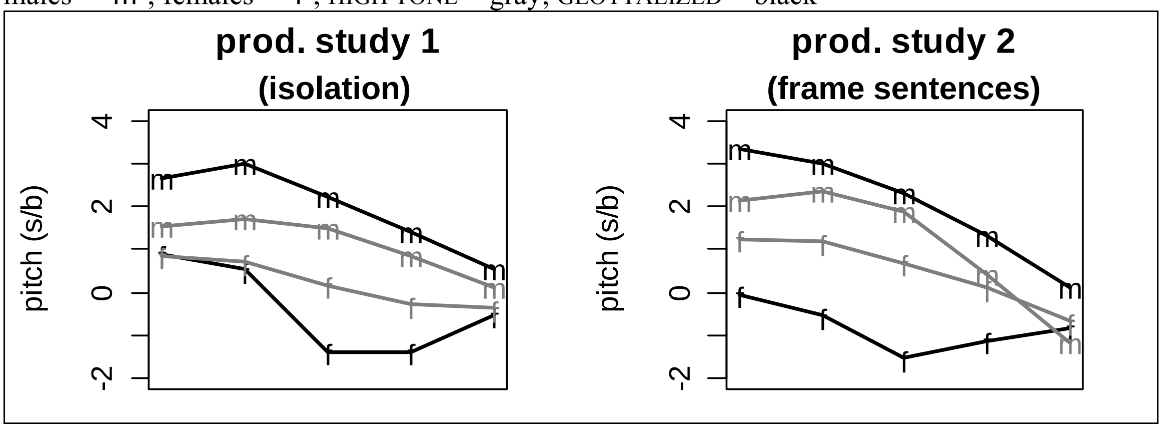
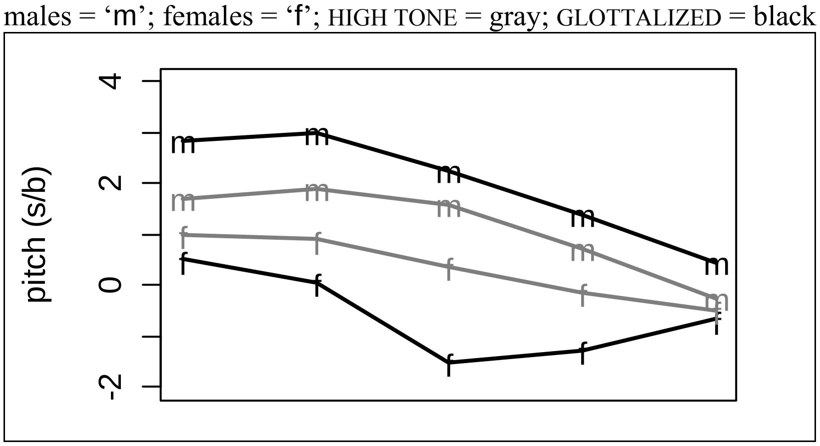
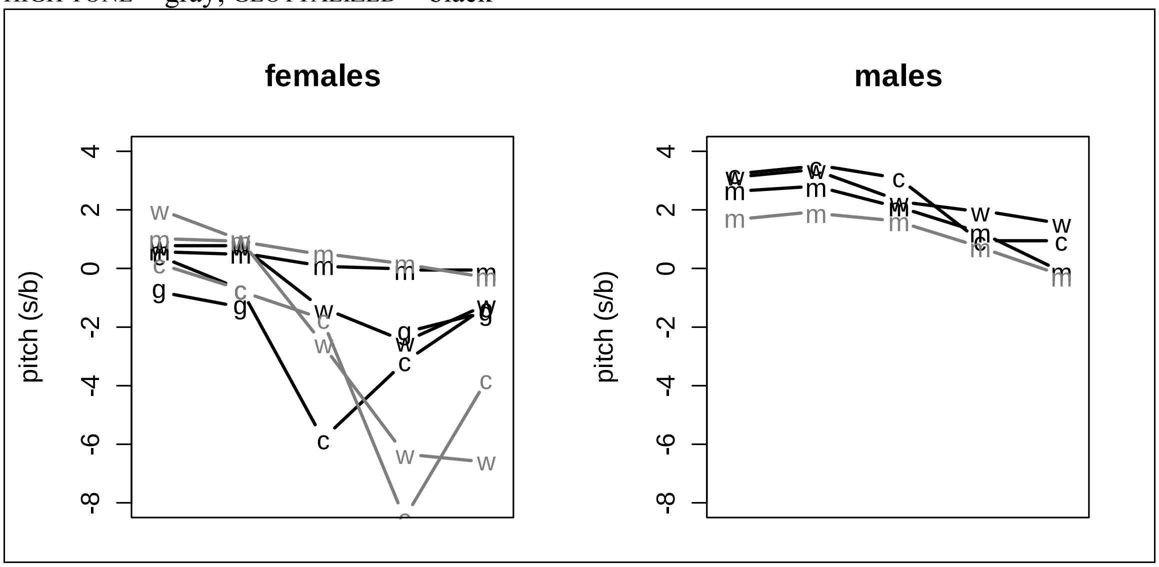


Figure 5.2: Average pitch contours for production studies 1 and 2 combined
males = 'm'; females = 'f'; HIGH TONE = gray; GLOTTALIZED = black



In order to better understand the interaction of pitch and creaky voice, Fig. 5.3 presents the average pitch contours for both HIGH TONE and GLOTTALIZED vowels by glottalization type and gender. What we see in Fig. 5.3 is that glottalization has a great effect on pitch contours as spoken by females but a minimal effect on pitch contours as spoken by males. Specifically, weak glottalization and creaky voice causes a dramatic dip in the middle or last portion of the pitch contour for females for productions of both HIGH TONE and GLOTTALIZED vowels.

Figure 5.3: Review of pitch by glottalization type (production studies 1 and 2 combined)
modal voice = ‘m’; weak glottalization = ‘w’; creaky voice = ‘c’; glottal stop = ‘g’;
HIGH TONE = gray; GLOTTALIZED = black



If we review Figs. 5.1-5.3 with the intention of determining how pitch is used to signal a contrast between HIGH TONE and GLOTTALIZED vowels, a problem arises, such that it appears males and females employ different strategies in the production of pitch. For males, GLOTTALIZED vowels always start with higher pitch than HIGH TONE vowels, regardless of the type of glottalization produced with the GLOTTALIZED vowels, but for females, the initial pitch of the vowel varies greatly for both vowel shapes.

One way to account for this gender-based difference would be to use a grammar that contains gender-specific constraints. Perhaps it is the case that the Yucatec Maya grammar tells males to produce GLOTTALIZED vowels with higher initial pitch than HIGH TONE vowels but does not tell females to do this. In the terms of our model of analysis, this could be handled with cue constraints that are sensitive to gender. The constraint */GLOTTALIZED/,[initial pitch = 4 s/b]$_{male}$ could have a different ranking value than */GLOTTALIZED/,[initial pitch = 4 s/b]$_{female}$ (for example), and so the optimality of the pairing of /GLOTTALIZED/ and [initial pitch = 4 s/b] would depend on gender. For the purposes of this dissertation, I see no reason to introduce such a complex innovation to the model and one empirical and one theoretical argument against it.[50]

According to the Bidirectional Model, cue constraints are used in both the production and perception grammar. This means that if gender-sensitive cue constraints are used to account for gender-based differences in the production data, then the perception grammar will be gender-sensitive as well. Depending on exactly how a constraint like */GLOTTALIZED/,[initial pitch = 4 s/b]$_{male}$ would assign violation marks (depending on if "male" is linked with the surface form, the phonetic form, or both), the use of such constraints would predict that different phonetic values for a given dimension are preferred by male listeners as opposed to female listeners or that different phonetic values as spoken by a male speaker are preferred as opposed to phonetic values spoken by a female speaker (or both are the case). However, as we saw in Chapter 3, the results of the perception experiments indicate that none of these scenarios occur with the Yucatec Maya listeners. Male and female listeners exhibit the same behavior with regard to how pitch influences perception, and the sex of the speaker does not influence the listener's behavior in this regard.

We could still attempt to use gender-sensitive *articulatory constraints*, and the

[50] It is clear from the sociolinguistic literature that language use may be influenced by gender in ways that are not explicable by biology alone. Foulkes (2006) provides a succinct overview of the research in this area. Such differences could be analyzed with gender-sensitive constraints. While I argue against the relevance of gender-sensitive constraints for the production of pitch in Yucatec Maya, I am not arguing against the general existence of such constraints.

empirical evidence presented above would not be a problem. Because articulatory constraints are not relevant to the perception grammar, the use of gender-sensitive articulatory constraints would not predict a gender-sensitive perception grammar. However, there are theoretical reasons to oppose this strategy. I believe that the gender-based differences in the production data can be explained by a physiological difference between males and females that interacts with gender-neutral constraints.

Note just how low the pitch values are for time points 3 and 4 when glottalization is produced by females – less than -6 s/b (which is half an octave below the baseline). In §2.2, we saw evidence that the pitch values produced during creaky voice are fixed such that if a speaker produces creaky voice there is an inherent pitch value (in Hertz) that will be produced. Thus, the speaker who uses creaky voice has no choice as to what pitch values to produce during that portion of the vowel.[51] If the grammar of Yucatec Maya wants GLOTTALIZED vowels to be produced with creaky voice during the middle of the vowel and with super high initial pitch (i.e. initial pitch that is higher than the initial pitch of HIGH TONE vowels), these two desires pose a problem for females that is not there for males. Namely, in order for females to satisfy both criteria, they would have to produce an initial pitch value of over 2 s/b and then drop to a pitch value of -6 s/b within the span of 100 ms. It seems likely

[51] The claim that the speaker "has no choice" could mean two different things. On one hand, it could mean that there is a physiological link between creaky voice and pitch that cannot be altered. On the other hand, it could also mean that the grammar of Yucatec Maya links pitch and creaky voice with such high-ranking constraints that the two dimensions cannot be controlled independently. In the context of the full Bidirectional Model, the first option would be encoded with sensorimotor constraints while the second would be handled with articulatory and/or cue constraints. The literature on other languages in which creaky voice and tone are used contrastively leads me to favor the first option, though more experimentation in this area is needed. Silverman (1997) claims that non-modal phonation and tone are always sequenced in Otomanguean languages and that this allows for the listener to recover information with respect to both tone and voice quality. For example, in Jalapa Mazatec (see also Silverman et al. 1995, Kirk et al. 1993), when tone occurs with either creaky or breathy voiced vowels, the non-modal phonation occurs during the initial portion of the vowel while the final portion of the vowel is nearly modal. I support Silverman in his assessment that this sequencing aids the listener, but it could also be the case that tone and non-modal phonation have to be sequenced in this way because pitch is linked to creaky voice. If the speaker is physically incapable of producing different tones while producing creaky voice, then such sequencing is the only way to maintain laryngeally complex contrasts.

that there are articulatory constraints that would prohibit a pitch change of such a magnitude over such a short duration. If these articulatory constraints are ranked above the constraints that require initial high pitch, the result would be that females are penalized for producing super high initial pitch whereas males are not. In this way, the gender differences fall out of the grammar without directly encoding them in the grammar. I believe that this strategy is preferable given that there is no evidence for a sociolinguistic usage of gender-differentiated pitch in Yucatec Maya.[52] There is no reason to assume the articulatory constraints themselves are gender-sensitive; it is the physiological reality that pitch during creaky voice is so far below the baseline for females that causes females to be susceptible to articulatory constraints that penalize large changes in fundamental frequency over short periods of time.

In order to better understand the claims just made, it will be useful to look at not just the pitch contours defined by time points 1-5 but also at the pitch span (the difference between the maximum and minimum pitch values produced for a given vowel). As described in §2.1.3.2, maximum and minimum pitch values for the duration of the vowel were obtained regardless of where in the vowel they were produced. This means that the pitch contours in Figs. 5.1-5.3 do not necessarily represent the extreme high and low pitch values of each contour. Table 5.2 provides the average values for initial pitch and pitch span by gender, vowel shape, and glottalization type. Averages are given for each production study separately and for both production studies combined. The patterns that emerge from these numbers support the claims made above.

[52] To be clear, no one has looked for such sociolinguistic functions of pitch and so this is an open area for research.

Table 5.2: Initial pitch and pitch span averages

			GLOTTALIZED			HIGH TONE		
			n	init. pitch	pitch span	n	init. pitch	pitch span
study 1	males	modal	54	2.4	3.7	95	1.5	3.8
		wg	27	3.1	5.6	0	-	-
		creaky	12	2.8	6.5	0	-	-
		gs	0	-	-	0	-	-
	females	modal	57	1.3	2.9	124	0.8	2.8
		wg	13	-0.2	6.2	4	2.0	9.4
		creaky	47	1.0	9.5	2	1.7	12.9
		gs	5	-1.3	6.5	0	-	-
study 2	males	modal	28	3.2	5.0	36	2.1	4.7
		wg	4	3.4	5.0	0	-	-
		creaky	4	4.4	7.8	0	-	-
		gs	0	-	-	0	-	-
	females	modal	51	-0.2	4.6	84	1.3	3.8
		wg	8	2.3	6.0	0	-	-
		creaky	26	-0.6	8.5	3	-0.9	10.4
		gs	1	1.7	11.6	0	-	-
all	males	modal	82	2.7	4.1	131	1.7	4.1
		wg	31	3.1	5.5	0	-	-
		creaky	16	3.2	6.8	0	-	-
		gs	0	-	-	0	-	-
	females	modal	108	0.6	3.7	208	1.0	3.2
		wg	21	0.8	6.1	4	2.0	9.4
		creaky	73	0.4	9.1	5	0.2	11.4
		gs	6	-0.8	7.4	0	-	-

The statistically significant differences among pitch span values for different glottalization types are given in Table 5.3. Here we see that pitch span increases with weak glottalization and more so with creaky voice (as compared to the pitch span produced with modal voice). Because only a handful of tokens were produced with a full glottal stop, there are few significant results for this glottalization type, and it is hard to generalize about the effect of the full glottal stop on pitch span. One thing to keep in mind is that glottal stops are usually (but not always) followed by creaky voice in this language, and so the size of a pitch span in a production with a full glottal stop may be dependant on whether or not the glottal stop is accompanied by creaky voice.

Table 5.3: Statistically significant differences among pitch span values
(p <.05, using a mixed linear regression model to account for multiple observations within subjects)

		GLOTTALIZED	HIGH TONE
study 1	males	cr., wg. > mod.	*n/a*
	females	cr. > wg. > mod.; cr. > gs.	cr., wg. > mod.
study 2	males	*no sig. differences*	*n/a*
	females	cr. > mod.	cr. > mod.
all	males	cr., wg. > mod.	*n/a*
	females	cr. > wg. > mod.	cr., wg. > mod.

The production of pitch spans is not significantly affected by gender. This supports the idea that females are limited as to how high pitch can be before creaky voice because of constraints on how big a pitch span can be. Furthermore, pitch span is not significantly affected by phonological vowel shape (within each category of glottalization type). On average, GLOTTALIZED vowels will have a larger pitch span because they are more commonly produced with glottalization, but there is no evidence at this time that a large pitch span is conditioned by the underlying category "GLOTTALIZED vowel".

When modal voice is produced, there is a tendency for GLOTTALIZED vowels to have a higher initial pitch than HIGH TONE vowels, but this tendency is not absolute (as demonstrated by females in production study 2). There are few significant results for the effect of vowel shape on initial pitch, as shown in Table 5.4.[53] However, of note is that when we combine both production studies, males produce a higher initial pitch for the GLOTTALIZED vowels, while females do not. This again reinforces the idea that high pitch is the goal in the production of the glottalized vowels, but this goal is not reached by females, who must produce very low pitch (as compared to their baseline) during creaky voice.

[53] One reason for the lack of significant results is how small the sample sizes are for many of the categories.

Table 5.4: Statistically significant differences among initial pitch values
(p <.05 using a mixed linear regression model to account for multiple observations within subjects)

study 1	males	*none*
	females	wg.: HI > GL
study 2	males	*none*
	females	mod.: HI > GL
all	males	mod.: GL > HI
	females	*none*

The above discussion has pointed to a number of interacting factors that could be used to signify the contrast between GLOTTALIZED and HIGH TONE vowels. GLOTTALIZED vowels tend to have a higher initial pitch and a larger pitch span, but these dimensions vary with glottalization type and gender. If, for example, HIGH TONE vowels are produced with weak glottalization or creaky voice, they have just as large of a pitch span as GLOTTALIZED vowels produced with weak glottalization or creaky voice. However, if we look at just those vowels produced with modal voice, we see that GLOTTALIZED vowels tend to have a larger pitch span than HIGH TONE vowels (though none of the differences are statistically significant, possibly due to small sample sizes). Thus, the pitch span of a vowel is directly correlated with glottalization type but is also at least indirectly correlated with vowel shape. Initial pitch, on the other hand is linked to glottalization type in a different way. While males produced GLOTTALIZED vowels with a higher initial pitch than HIGH TONE vowels, females did not. I have proposed that this is because articulatory constraints penalize the production of a large pitch spans. This means that when females produce creaky voice, they are limited as to how high their initial pitch can be. We see strong evidence of this in production study 1, where females produce those GLOTTALIZED vowels with modal voice with a higher initial pitch value than both HIGH TONE vowels with modal voice and GLOTTALIZED vowels with some form of glottalization. The phonetic dimensions that seem to indicate a contrast between GLOTTALIZED and HIGH TONE vowels are summarized in Table 5.5, and the goal of the rest of this chapter is to determine how best to design the grammar so that all of these dimensions are accounted for.

Table 5.5: Phonetic dimensions that distinguish GLOTTALIZED and HIGH TONE vowels

	GLOTTALIZED	HIGH TONE
glottalization	more weak glottalization, creaky voice, and glottal stop	more modal voice
initial pitch	higher (with modal voice)	lower (with modal voice)
pitch difference	larger (especially with glottalization)	smaller

5.1.2 Perception Data

As described in Chapter 3, 14 participants from Santa Elena participated in a perception study that involved two tasks. Because the results of the tasks are so different and because task 1 involved natural stimuli (whereas task 2 involved manipulated stimuli), only the results of task 1 will be considered in evaluating the perception grammar of native language users from Santa Elena.[54] The results of this task are summarized in Table 5.6, where we see the percentage of times that participants chose a GLOTTALIZED vowel (a HIGH TONE vowel was chosen every other time) on the basis of stimulus glottalization type, initial pitch, and pitch span.

Table 5.6: Review of results for task 1 (Santa Elena only)

Percentage of times a GLOTTALIZED vowel was chosen on the basis of stimulus glottalization type, initial pitch, and pitch span.

glottalization type					
ng	wg	cr	gs		
39	68	55	93		
initial pitch category					
-2	0	2	4	6	8
52	44	38	59	79	50
pitch span category					
2	4	6	8	10	12
33	41	59	63	71	84

[54] The implications of the different results from the two tasks were discussed in §3.5.

The generalizations that we can make from this table are that more glottalization and a larger pitch span will trigger more GLOTTALIZED vowel responses. The results for initial pitch are not so straightforward. There is a significant effect of initial pitch in that as initial pitch increases, participants are more likely to select a GLOTTALIZED vowel, but this relationship is not uniform as it is with pitch span. Notably, the lowest initial pitch category of [-2] triggers more GLOTTALIZED vowel responses than the higher categories of [0] and [2], while the highest category of [8] is evenly split between HIGH TONE and GLOTTALIZED vowel responses.

When we compare the results of task 1 to the production data presented above, there are many symmetries. The acoustic cues that can be used to group the production data are all used by the listener in a discrimination task. In the rest of this chapter, we will work on how to account for these symmetries with the Bidirectional Model and a GLA learner.

5.2 Constraints

There are three types of constraints that are applicable to the analysis of the phonetics-phonology interface with the Bidirectional model: structural constraints, cue constraints, and articulatory constraints. Structural constraints, which penalize marked surface forms, are only relevant in the perception grammar. There are only two surface forms of interest here – GLOTTALIZED and HIGH TONE vowels – which are both legitimate structures in this language. The constraints that penalize them should all be lowly ranked, and so I will not use structural constraints in these analyses. In this section, I will first propose the cue constraints that can account for the generalizations about the data that were just made and then I will propose the relevant articulatory constraints. This section concludes with a discussion about how to use the mean ranking values of these constraints (as will be determined by the GLA) to generalize about the patterns in the production and perception of the language.

5.2.1 Cue Constraints

Cue constraints are proposed to be of the form */y/,[x] and to penalize any pairing of a surface form /y/ with a phonetic form [x]. In the production grammar, this means that */y/,[x] assigns a violation mark for each [x] in a phonetic form (a candidate) if and only if /y/ is in the surface form (the input). Thus the constraint penalizes a speaker who uses [x] to express the sentiment /y/. In the perception grammar, */y/,[x] assigns a violation mark for each /y/ in a surface form candidate if and only if [x] is in the phonetic input, and so this constraint penalizes a listener who takes [x] to mean /y/.

In order to define cue constraints for our current grammar, we first need to determine the relevant values for /y/ and [x]. Considering first the possible values for /y/, there are only two surface forms of interest for our grammar: the surface form of HIGH TONE vowels and the surface form of GLOTTALIZED vowels. In Table 2.21, I proposed that these phonological forms were /v́v/ and /v́v̰/, respectively. These proposals were based on the phonetic evidence presented in Chapter 2 and encoded information on length, tone, and glottalization. For the purposes of learning this grammar, though, it is preferable to work with more loosely defined surface forms. The goal of this section is to use a simulated learner to develop the rankings of cue constraints that best account for the phonetic data. The grammar that is the outcome of a learning simulation will then show us, by the ranking of the cue constraints, which phonetic dimensions are successful at distinguishing HIGH TONE vowels from GLOTTALIZED vowels. Thus, for the initial stage of our learning simulations, I assume that the learner knows there are two distinct categories, which I call HIGH TONE and GLOTTALIZED, but which could also be called CATEGORY A and CATEGORY B, or any other labels.[55] I will henceforth abbreviate the two possible inputs to our production grammar as /gl/ and /hi/.

The relevant values for the phonetic forms [x] require more consideration. We need to define both the phonetic dimensions of interest and how to categorize each dimension. Because phonetic forms are continuous, the ideal system is one in which a cue constraint refers to a very finely grained category of some dimension. For example, if the dimension is pitch, we may want to categorize pitch values by a tenth or a hundredth of a semitone, whereas we certainly don't want categories that are 4 semitones wide. The smaller the categories of a given dimension, the more data we need to develop a meaningful grammar. Because I am working with a data set of only 685 tokens, it will not be possible to use cue constraints that reference categories that are as small as we would like. This is illustrated below with the dimension of initial pitch.

If I categorize initial pitch by one tenth of a semitone and consider the full range of the data for GLOTTALIZED vowels, we see in Fig. 5.4 that, while this data roughly follows a normal distribution, there are many inconsistencies due to the sparseness of the data points compared to so many categories. Surely it is simply a fact about this data set that -10.4 s/b is produced three times and -10.3 s/b never; some other data set obtained from the same group of speakers performing the same task could have one or more tokens with -10.3 s/b and none with -10.4 s/b. In Fig. 5.5, on the other hand, wider categories of 1 and 2 semitones are used to categorize the data, and the idiosyncrasies are mostly obscured.

[55] Boersma et al. (2003) describe how the learner first uses *auditory-driven learning* to learn the discrete phonological categories of the language. These categories are then given arbitrary labels. This is the stage that our hypothetical learner is at.

Figure 5.4: Distribution of initial pitch values for GLOTTALIZED vowels as categorized by 1/10 semitones

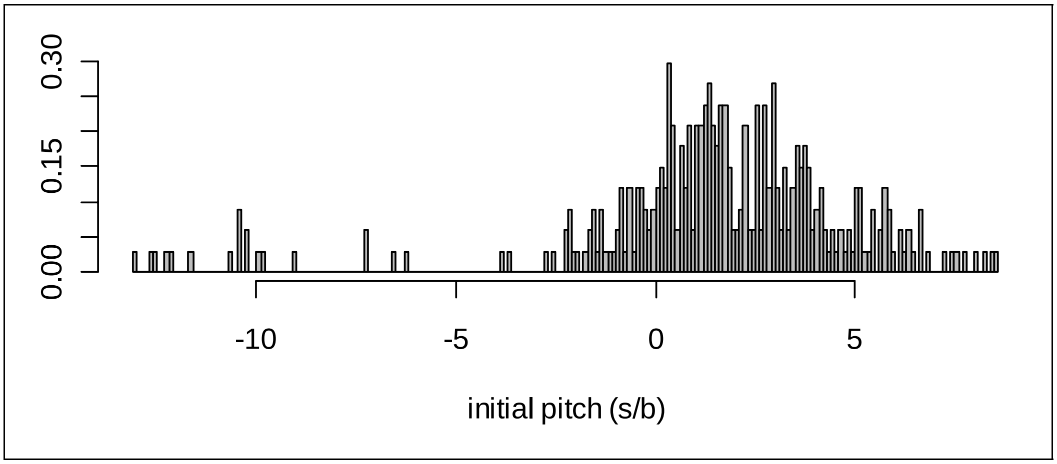
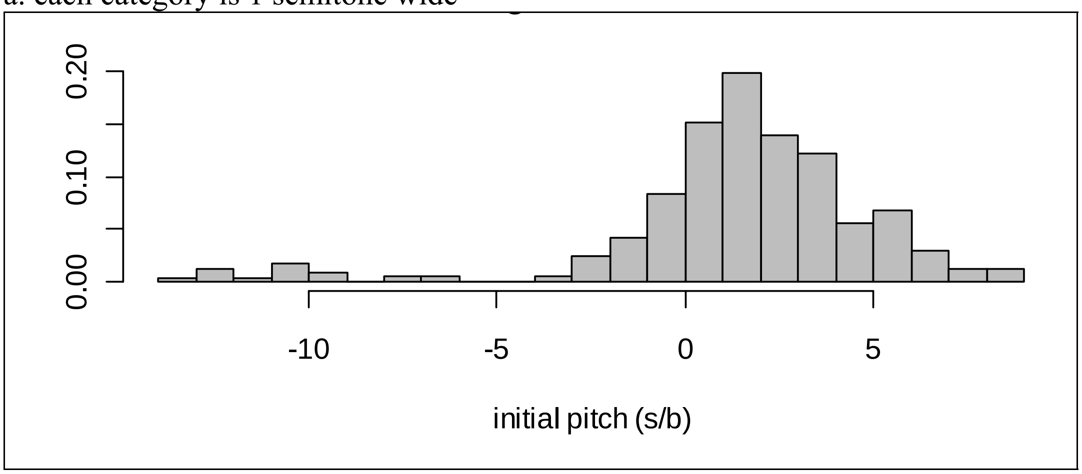
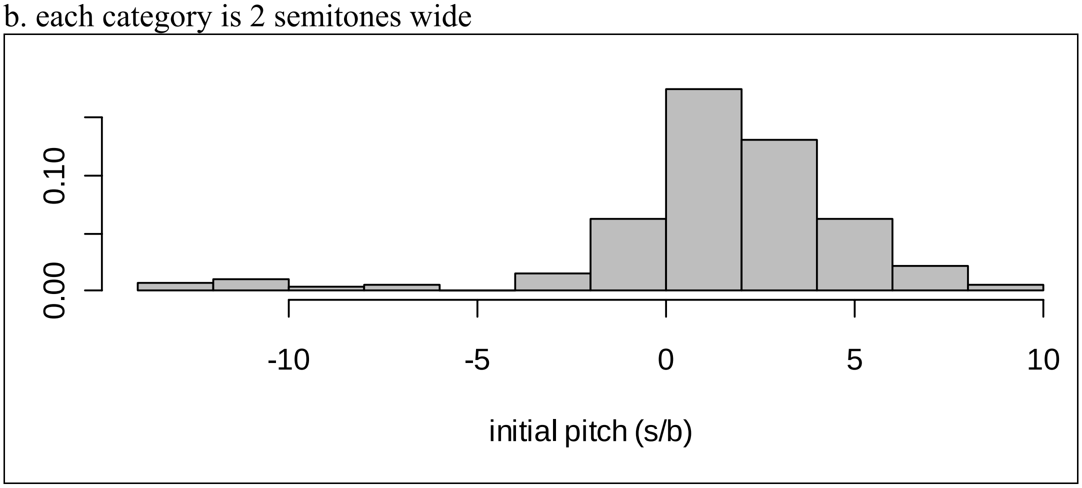


Figure 5.5: Distribution of initial pitch values for GLOTTALIZED vowels as categorized by wider categories

a. each category is 1 semitone wide



b. each category is 2 semitones wide



Another concern in determining how finely to categorize a given dimension is that we want the grammar to be intelligible to the linguist. As we add more and more dimensions to the model with more finely grained cue constraints, we get a grammar that is more and more complex. While this complex grammar is our ultimate goal in modeling the human language user, it also obscures important generalizations that we wish to make as linguists.

Thus, the size of the data set places limitations on how finely grained the constraints can be for the development of a meaningful analysis. Furthermore, the smaller the constraint set the easier it is to make generalizations about the grammar. For these reasons, I define a grammar that uses wider categories than would be ideal for analysis of the phonetics-phonology interface. We will use this grammar to make important generalizations about the data and the model. In §6.1.1 I will illustrate how some interesting patterns do emerge when more finely grained constraints are used even with such a sparse data set. This will show that work in this area with more appropriate constraints is promising, but does require a large body of data to work from. In the following discussion, I will identify what I consider to be the ideal way to categorize some phonetic dimension and then define how that phonetic dimension will be categorized for this grammar as directed by the limitations of the number of data points.

There are two types of pitch values that we want to consider regulating with cue constraints: initial pitch, and pitch span. In determining how to categorize initial pitch, we must balance the competing goals of using finely grained constraints and using constraints that are appropriate for the number of tokens in the data set. After running a variety of analyses with constraints defined by categories of different widths, I have found that dividing pitch into categories of two semitones wide leads to a grammar that both captures the important patterns and is able to be displayed in text in an understandable way.

I will also use two semitones as the division of pitch span measurements for creating cue constraints that regulate pitch span. With regard to pitch spans, we may question whether or not cue constraints are appropriate. Recall that this measurement is correlated with vowel shape, but indirectly; because glottalization is correlated with both vowel shape and pitch span, there appears to be a correlation between vowel shape and pitch span. In other words, vowel shape does not account for variation in pitch span above and beyond the variation that is accounted for by glottalization type. This suggests that perhaps cue constraints should not be used to link vowel shape with pitch span but that instead articulatory constraints might dictate what the preferred pitch spans are. For example, an articulatory constraint might penalize the production of creaky voice and a small pitch span in the same phonetic form. However, while it may be the case that the production grammar does not necessarily assign pitch spans on the basis of phonological vowel shape, the Bidirectional Model tells us that cue constraints are also relevant in the perception grammar. Since vowel shape is correlated with pitch span, regardless of the cause of that correlation, it is possible that the listener has not only learned this correlation but uses it in perception. Thus, even if the original motivation for the production of a large pitch span is articulatory, if the listener uses this cue, then we must have cue constraints that regulate this cue in the model. This point illustrates how both production and perception concerns are encoded in both sides of the grammar with the Bidirectional Model. Of course, the results of the perception experiments that were presented in Chapter 3 indicate that listeners do use pitch span as a cue to distinguishing HIGH TONE from GLOTTALIZED vowels, and so there is a reason for including these cue constraints in our model.

The categorization of initial pitch and pitch span values is summarized in (5.1). Each category shown is hence one possible value for [x], e.g. a certain phonetic form may be marked as [initial pitch = 0] or [pitch span = 4], etc. For simplicity, I do not use categories that are two semitones wide for the entire range of produced initial pitch and pitch span values. Because 90% of initial pitch values are between -2 and 6 s/b, I put all initial pitch values below -2 s/b in one category and all those above 6 s/b in another. Similarly, only 6% of pitch span measurements are above 12 semitones, and so I put all these pitch span values in one category.

(5.1) categories defined for the phonetic dimensions of initial pitch and pitch span

initial pitch = x		pitch span = x	
[-2]	$x \leq -2$	[2]	$x \leq 2$
[0]	$-2 < x \leq 0$	[4]	$2 < x \leq 4$
[2]	$0 < x \leq 2$	[6]	$4 < x \leq 6$
[4]	$2 < x \leq 4$	[8]	$6 < x \leq 8$
[6]	$4 < x \leq 6$	[10]	$8 < x \leq 10$
[8]	$6 < x$	[12]	$10 < x$

There is one more important factor to consider in defining the phonetic forms ([x]). It is clear that glottalization type is a phonetic dimension that should be controlled by cue constraints. The relevant categories of this dimension are minimally *modal voice*, *weak glottalization*, *creaky voice*, and *glottal stop* (see §2.1.3.1). Time is also relevant: modal voice and non-modal voice are preferred at different time points during vowel production. Ideally, we would want our values for this dimension to couple different time points with each of the four glottalization types, i.e. [modal, beg. of vowel], [modal, middle of vowel], [modal, end of vowel], etc. Just how many time points to reference (i.e. how finely to categorize vowel duration) is an open question. For simplicity, I do not encode time with these constraints, and so we will use the four main categories of glottalization, as shown in

(5.2).

(5.2) categories defined for the phonetic dimension of glottalization

[modal]	modal voice is produced throughout vowel production
[weak glottalization]	weak glottalization is produced at some point of vowel production
[creaky]	creaky voice is produced at some point of vowel production
[glottal stop][56]	a glottal stop is produced at some point of vowel production

We have now identified the relevant values for /y/ (/gl/ and /hi/) and for [x] (various categories of initial pitch, pitch span, and glottalization type), and so we now need to use these values to define cue constraints. Because our surface forms (/gl/ and /hi/) are simply labels for two different phonological categories, there is no reason to limit which combinations of the possible values for /y/ and [x] should be penalized by cue constraints (Escudero and Boersma 2004). In other words, any arbitrary phonological category could, in theory, be paired with any acoustic value on any dimension.[57] Thus, I will define a cue constraint by pairing each value for /y/ with each value for [x]. Table 5.7 lists all the cue constraints that will be used in the analysis of the next section. These constraints will be consistently abbreviated as they are in this table.

[56] Yucatec Maya uses a phonemic consonantal glottal stop both in lexical representations (e.g. *na'* |naʔ| 'grandmother') and as an epenthetic consonant (e.g. *amigo* /ʔáamigóoh/ 'friend') (see §1.2). I assume that the phonetic implementation of a consonantal glottal stop ([glottal stop]$_C$) is controlled by different constraints than those that refer to the production of a glottal stop that interrupts a single vowel sound or syllable nucleus ([glottal stop]$_V$). Because consonants are not analyzed here, I use the phonetic form [glottal stop] as shorthand for [glottal stop]$_V$.

[57] If the category labels were not arbitrary, on the other hand, there are some pairings of /y/ and [x] values that would form suspect cue constraints. For example, if the surface form /gl/ was taken to mean 'marked for high pitch and creaky voice' (i.e. /v̰́/), the constraint */gl/,[cr] would basically be saying 'don't produce something marked for creaky voice with creaky voice' and would thus be an unnatural constraint. I discuss some of the implications of this in §6.1.3.

Table 5.7: Cue constraints

glottalization	*/gl/,[mod]	*/hi/,[mod]
	*/gl/,[wg]	*/hi/,[wg]
	*/gl/,[cr]	*/hi/,[cr]
	*/gl/,[gs]	*/hi/,[gs]
initial pitch	*/gl/,[in = -2]	*/hi/,[in = -2]
	*/gl/,[in = 0]	*/hi/,[in = 0]
	*/gl/,[in = 2]	*/hi/,[in = 2]
	*/gl/,[in = 4]	*/hi/,[in = 4]
	*/gl/,[in = 6]	*/hi/,[in = 6]
	*/gl/,[in = 8]	*/hi/,[in = 8]
pitch span	*/gl/,[sp = 2]	*/hi/,[sp = 2]
	*/gl/,[sp = 4]	*/hi/,[sp = 4]
	*/gl/,[sp = 6]	*/hi/,[sp = 6]
	*/gl/,[sp = 8]	*/hi/,[sp = 8]
	*/gl/,[sp = 10]	*/hi/,[sp = 10]
	*/gl/,[sp = 12]	*/hi/,[sp = 12]

5.2.2 Articulatory Constraints

Articulatory constraints penalize effortful productions at the phonetic level. In this section, I propose phonetically plausible articulatory constraints that are related to the same phonetic dimensions as those regulated by cue constraints. Following de Lacy (2002) and Prince (1997a-b), I assume that articulatory constraints are scalar in that they penalize a particular production and any other productions that are more effortful on the same dimension.[58] For example, if a constraint penalizes the production of a pitch span of [8] because it is too effortful to produce such a change in vocal fold vibration, then this constraint should also penalize any pitch span larger than [8] because these productions are even more effortful (within the given phonetic dimension). Articulatory constraints only look at phonetic forms when they are output candidates and so they are only relevant in the

[58] De Lacy's theory of markedness (which was developed in a phonological context and not in terms of the phonetics-phonology interface) also requires a set of scalar faithfulness constraints and no fixed rankings. The cue constraints proposed above do not conform to this aspect of the theory; it is an open question as to what the consequences of such cue constraints would be in the context of Bidirectional StOT. Fixed rankings are discussed in §5.7. Additionally, de Lacy requires that every point on some scale be referred to by some markedness constraint. I follow Gouskova (2003) in using a system where certain 'unmarked' values are not referred to by articulatory constraints.

production grammar.

As was discussed in §4.3.2, articulatory constraints actually look at the [articulatory form], which in our model is conflated with the [auditory form]. Thus, an articulatory constraint like *[pitch span = 8] really says "Do not change the rate of vocal fold vibration by greater than or equal to 8 semitones". Because the model assumes a universal link between an articulatory form with such a change in rate of vocal fold vibration and an auditory form with [pitch span = 8], this articulatory constraint also says "Do not produce an auditory form with [pitch span = 8]". In (5.3) below, I define the articulatory constraints used in this section with both types of definitions, but for consistency with the cue constraints, I will use the auditory-based definition throughout this paper.

With regard to glottalization, it is likely that there are articulatory constraints that penalize tight adduction of the vocal folds,[59] and hence the production of a full glottal stop, the production of creaky voice, and even the production of weak glottalization. I will thus use a constraint *[wg] that penalizes [weak glottalization], [creaky voice], and a [glottal stop], a constraint *[cr] that penalizes [creaky voice] and a [glottal stop], and a constraint *[gs] that penalizes only a [glottal stop]. (See (5.3) below for full definitions of all markedness constraints.)

There are cross-linguistic reasons to propose cue constraints that penalize either high or low pitch values, as these constraints would be likely to do work in non-tone languages. However, with respect to the data under consideration here, such constraints will be ineffective because high pitch values are prevalent in the data (hence articulatory constraints

[59] Based on Edmondson and Esling (2006), it is possible that articulatory constraints penalize the use of the different valves of the throat, and so, for example, we might want to penalize the engagement of Valves 1 and 3 (either with one constraint or two independently rankable constraints). I have chosen to simply reference vocal fold adduction because I am not sure exactly what laryngeal maneuvers are involved with "weak glottalization".

penalizing them would end up with extremely low ranking values) and low pitch values are almost nonexistent (and hence articulatory constraints penalizing them would end up with extremely high ranking values).[60] Thus, I will not use constraints that penalize certain values for initial pitch in this model.

There are two types of articulatory constraints with respect to pitch spans that are appropriate for our purposes. As discussed above, there is a preference for producing a large pitch span in conjunction with creaky voice. I propose that this preference is (at least in part) conditioned by articulatory constraints that penalize the production of creaky voice and a small pitch span in the same phonetic form. I do not yet have evidence that the production of creaky voice in combination with a small pitch span is effortful, but there is a cross-linguistic correlation between creaky voice and preceding high pitch (which equates with a large pitch span (at least for females) because creaky voice is produced with lower pitch). Acoma (a Keresan language of New Mexico) has a 'glottal accent', which is described in a manner similar to descriptions of Yucatec Maya's GLOTTALIZED vowel; it is produced with a falling pitch contour and creaky voice (Miller 1965). The Danish *stød* begins with high pitch and ends with creaky voice (Fischer-Jørgensen 1989), and vowels before glottalized sonorants are produced with high pitch in Coatlán-Loxicha Zapotec (Plauché et al. 1998). The correlation is thus found synchronically in multiple language families, and diachronically it is widely believed that a coda glottal stop (which is likely correlated with creaky voice in the vowel) can condition a rising pitch contour in the preceding vowel (Hombert 1978). Based on this evidence, I propose articulatory constraints that penalize the pairing of glottalization ([creaky

[60] If we looked at all vowel shapes in Yucatec Maya, articulatory constraints penalizing both high and low pitch values would be useful for maintaining the mid pitch values associated with the SHORT vowel. In this way, constraints that penalize high or low pitch values can also do work in tone language in that they can determine what the unmarked pitch values are.

voice] or [weak glottalization]) with lower pitch span values in the phonetic form.

I also explained above that there should be constraints that penalize too large of a pitch span and that these constraints could be used to explain why females do not produce as high of an initial pitch as males for the GLOTTALIZED vowel. I will thus also use articulatory constraints that penalize large pitch spans, regardless of any other content in the phonetic form (i.e. if a pitch span is too large it will incur a violation mark regardless of whether it's produced with creaky or modal voice). As with the cue constraints, articulatory constraints must reference properties of continuous forms, and for this reason they should refer to finer-grained categories than the markedness constraints of phonological analysis. I will use the same divisions of pitch values that were defined above for the cue constraints. However, not all values are appropriate for use with articulatory constraints. For example, there is no motivation for a constraint *[sp=0] (don't produce a pitch span of 0 semitones) because such a production would not be effortful or cross-linguistically rare. Similarly, I do not propose articulatory constraints that penalize [modal voice], [sp=2], or [creaky voice, sp=8].

(5.3) articulatory constraints

<u>constraints that penalize glottalization</u>

*[wg]	Do not tightly adduct the vocal folds at any point of vowel articulation. =Do not produce a phonetic form which contains the auditory correlate of tightly adducted vocal folds at any point.
*[cr]	Do not tightly adduct the vocal folds for more than a few pitch periods during vowel production. =Do not produce a phonetic form which contains the auditory correlate of tightly adducted vocal folds for more than a few pitch periods.
*[gs]	Do not completely adduct the vocal folds during vowel production. =Do not produce a [glottal stop] in the phonetic form.

<u>constraints that penalize the cooccurance of glottalization and a small pitch span</u>

*[cr, sp ≤ 2]	If producing tightly adducted vocal folds, do not change the rate of vocal fold vibration by less than or equal to 2 semitones. =Do not produce a phonetic form with glottalization ([creaky voice] or [weak glottalization]) and a pitch span ≤ [2].
*[cr, sp ≤ 4]	If producing tightly adducted vocal folds, do change the rate of vocal fold vibration by e less than or equal to 4 semitones. =Do not produce a phonetic form with glottalization ([creaky voice] or [weak glottalization]) and a pitch span ≤ [4].
*[cr, sp ≤ 6]	If producing tightly adducted vocal folds, do change the rate of vocal fold vibration by less than or equal to 6 semitones. =Do not produce a phonetic form with glottalization ([creaky voice] or [weak glottalization]) and a pitch span ≤ [6].

<u>constraints that penalize large pitch spans</u>

*[sp ≥ 4]	Do not change the rate of vocal fold vibration by greater than or equal to 4 semitones. =Do not produce a pitch span ≥ [4].
*[sp ≥ 6]	Do not change the rate of vocal fold vibration by greater than or equal to 6 semitones. =Do not produce a pitch span ≥ [6].
*[sp ≥ 8]	Do not change the rate of vocal fold vibration by greater than or equal to 8 semitones. =Do not produce a pitch span ≥ [8].
*[sp ≥ 10]	Do not change the rate of vocal fold vibration by greater than or equal to 10 semitones. =Do not produce a pitch span ≥ [10].
*[sp ≥ 12]	Do not change the rate of vocal fold vibration by greater than or equal to 12 semitones. =Do not produce a pitch span ≥ [12].

5.2.3 Interpreting Mean Ranking Values

This section is primarily designed for the reader who is not familiar with Boersma (and company)'s writings on the mean ranking values of cue constraints within the context of StOT, but I will also try to make some ideas explicit that I believe have only been implied in previous discussions of the model. The goal of this section is to elucidate the types of patterns that are demonstrated by the mean ranking values of certain sets of constraints. Even for those who do research in classic OT and who can readily discern the predictions made by the dominance relations among a group of constraints, the meaning of the mean ranking values of a set of constraints in Stochastic OT is not necessarily obvious. In this section I will discuss the different sets of constraints that are relevant for each type of evaluation (production and perception) and how the relations among the mean ranking values of the constraints in a given set allow us to make generalizations about the grammar.

For any point of evaluation (i.e. a tableau), the optimal output candidate (from a set of candidates) is determined with reference to a given input and a given constraint ranking. It is not necessary to consider all the constraints of the grammar at every point of evaluation; only those constraints that conflict, that are capable of assigning different numbers of violation marks to the members of the candidate set, need to be considered. A trivial example is the fact that articulatory constraints are not relevant in the perception grammar. In this grammar, an auditory form is the input and surface forms compete as candidates. Articulatory constraints cannot assign violation marks to surface forms, and so these constraints would never be able to assign a fatal violation mark to any of the candidates. For this reason, it would be meaningless to even consider the ranking values of articulatory constraints when showing a point of analysis in the task of perception/comprehension.

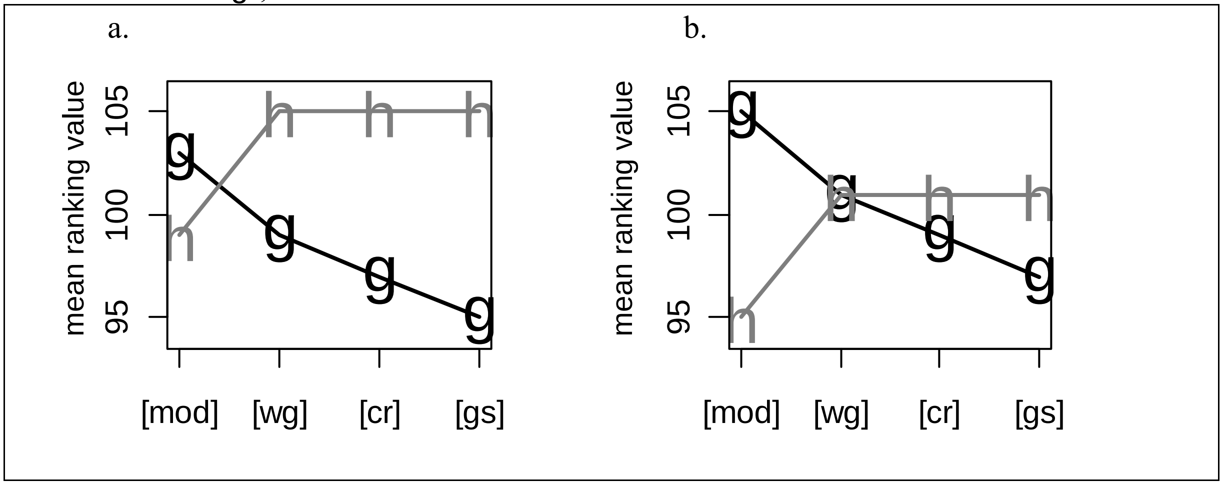
The cue constraints are relevant to both the production and perception grammars, but only a subset of these constraints are relevant for any single point of evaluation. If we use our Yucatec Maya production grammar to evaluate the input /gl/, none of the cue constraints that refer to /hi/ will be relevant because none of them will assign violation marks to any candidates. To generalize, when considering the production grammar, only and all the cue constraints that refer to /y/ will be relevant when /y/ is the input. If /gl/ is the input, a tableau should consider all the cue constraints in the "/gl/ set" – the set of cue constraints that penalize the pairings of certain acoustic values with /gl/ – and none of the constraints in the "/hi/ set".

In perception, phonetic forms are the inputs. This means that all and only those cue constraints that refer to a given [x] will be relevant when [x] is the input. If [modal voice] is (part of) the input, the "[modal voice] set" of cue constraints is relevant, while the "[creaky voice]" set is not.[61]

Once we have determined mean ranking values of the relevant constraints, we can make generalizations about the production and perception grammars by focusing on the set of constraints that refers to the values for /y/ and [x] that are of interest. For example, if we wish to look at the relationship between surface forms and glottalization type, we can focus on the constraint set {*/gl/,[mod]; */gl/,[wg]; */gl/,[cr]; */gl/,[gs]; */hi/,[mod]; */hi/,[wg]; */hi/,[cr]; */hi/,[gs]}. Two hypothetical rankings of this constraint set are shown in Fig. 5.6. This figure will allows us to familiarize ourselves with how constraint rankings will be displayed in this paper and how to use such graphs to make generalizations.

[61] The consequences of production and perception tableaus being associated with different sets of relevant constraints are addressed in §5.3.

Figure 5.6: Hypothetical ranking of cue constraints
GLOTTALIZED = 'g'; HIGH TONE = 'h'



In each of the graphs above, the ranking of some cue constraint is denoted by one point on a line. The cue constraint can be identified by the column, which is labeled with a value for [x], and the line, which is labeled with a value for /y/. Hence the left-most 'g' in Figure 5.6 tells us the mean ranking value of */gl/,[mod]. Because the cue constraints are negatively formulated, constraints with lower mean ranking values denote preferred surface form – phonetic form pairings.

If we wish to make generalizations about the production grammar, the constraints on a given line tell us the information we need to know. If look at just the 'g' line or just the 'h' line, they have the same shape in both graphs. Both graphs show that, in these hypothetical production grammars, HIGH TONE vowels will mostly be produced with modal voice. The constraint */hi/,[mod] has the lowest mean ranking value of all the constraints in the /hi/ set and so this constraint denotes the preferred value for [x] when /hi/ is the input to production. Similarly, the graphs also show that GLOTTALIZED vowels will tend to be produced with a glottal stop or (less often) creaky voice or (even less often) weak glottalization because */gl/,[gs] has the lowest mean ranking value (followed by */gl/,[cr] and then */gl/,[wg]).

Because all of these glottalization types are marked – there are articulatory constraints that penalize their productions – we would need to know the mean ranking values of these articulatory constraints before making assumptions about how often a glottal stop would actually be produced.

While the graphs tell the same story in terms of production, they show two very different perception grammars. When generalizing about the perception grammar, it is the constraints in a given column that compete. If the input to perception is a token with modal voice, both grammars show that this phonetic form is preferably mapped onto a HIGH TONE vowel, because the constraint */hi/,[mod] has a lower mean ranking value than */gl/,[mod], meaning that the pairing of /hi/ and [mod] is more preferable than the pairing of /gl/ and [mod]. However, only grammar (b) almost guarantees this mapping. Because of how far apart the ranking values of */hi/,[mod] and */gl/,[mod] are, it is highly unlikely that the evaluation ranking value of */hi/,[mod] will be higher than */gl/,[mod], which is the ranking needed in order for [mod] to be mapped onto /gl/ in perception.

If the input is a token with weak glottalization, on the other hand, this token will likely be mapped onto a GLOTTALIZED vowel with grammar (a), but there would be an almost even split between HIGH TONE and GLOTTALIZED mappings with grammar (b). Similarly, grammar (a) – more so than grammar (b) – predicts glottalization in general to be a strong cue to the perception of a GLOTTALIZED vowel.

We have now seen how the production and perception grammars are defined by different sets of constraints. Two identical production grammars are not necessarily associated with two identical perception grammars, and vice versa. In the types of graphs that are used in this dissertation, constraints on a given line conflict in production, while constraints in a given column conflict in perception.

5.3 Lexicon-Driven Perceptual Learning and Other Learning Strategies

In §5.2 we determined the relevant values for /y/ and [x], how these values are paired in order to define cue constraints, and the relevant articulatory constraints for our model. We have yet to develop a learning strategy. The GLA defines how and when constraint ranking values are adjusted by the learner, but there are various ways for the learner to use the data of the linguistic environment to test the interim grammar.

As discussed above, different sets of cue constraints are relevant for production tableaus than those that are relevant for perception tableaus.[62] In a production tableau, where a surface form /y/ is the input, the cue constraints that refer to /y/ are relevant. This means that if a Yucatec Maya GLA learner is trained with production tableaus, they can learn the relations among the cue constraints in the /gl/ set and the relations among the cue constraints in the /hi/ set, but they cannot learn the relation between a constraint from the /gl/ set (e.g. */gl/,[modal voice]) and a constraint from the /hi/ set (e.g. */hi/,[modal voice]). This last relation can be learned with perception tableaus, where [x] is the input. If perception tableaus are used, the GLA learner can learn the relations among the cue constraints in the [modal voice] set (etc.), but they cannot learn the relations among the /gl/ set or the /hi/ set.

Consider a scenario where a language learner hears someone say *k'a'an* (/gl/) with the phonetic form [creaky voice, in = 4, sp = 6]. There are two different ways for the learner to test their grammar: the learner may check to see if the current production grammar predicts the input /gl/ to be mapped onto [creaky voice, in = 4, sp = 6] and/or the learner may check to see if the current perception grammar predicts the phonetic form [creaky voice, in =

[62] Articulatory constraints are of course only relevant in production tableaus. The mean ranking values of articulatory constraints (and whether or not they are learned, and if so how) are discussed later in this section and again in §§5.5 and 5.7.

4, sp = 6] to be mapped onto the surface form /gl/. Both tests of the current grammar could potentially yield different results (e.g. the production grammar may predict the right output while the perception grammar may not) and thus could lead to different ranking adjustments.

It has been proposed in the literature that perception learning happens first (e.g. Pater 2004). Empirical evidence supports this claim: children can correctly perceive adult forms before they can produce them. Hence, Boersma (2006), Escudero and Boersma (2004), and Boersma and Hamann (2008) propose that ranking values of cue constraints are learned through *lexicon-driven perceptual learning*.[63] This means that the language learner takes the current training datum (the phonetic form just produced by some speaker) and checks to see if the interim perception grammar predicts that phonetic form to be mapped onto the correct surface form (that the speaker intended). If the prediction is correct, no ranking adjustment is needed, but if the prediction is incorrect, as shown in (5.4), then the mean ranking values of cue constraints are adjusted in accord with the GLA.

(5.4) perception learning

The learner hears [creaky voice, in =4, sp = 6] as a production for /gl/ (denoted by ✋). The current perception grammar predicts /hi/ to win (denoted by ☞), and so a ranking adjustment occurs, as denoted by the arrows.

	→	←	←	←	→	→
[creaky voice, in =4, sp = 6]	*/gl/, [in=4]	*/hi/, [cr]	*/hi/, [in=4]	*/hi/, [sp=6]	*/gl/, [sp=6]	*/gl/, [cr]
☞ /hi/		*	*	*		
✋ /gl/	*!				*	*

In the rest of this dissertation, I refer to any situation where the learner tests the interim grammar with perception tableaus as *perception learning* and to any situation where the learner tests the interim grammar with production tableaus as *production learning*.

[63] This type of learning is "lexicon-driven" because it makes use of the learner's knowledge about the lexical item that the speaker said. In other words, the learner must know that (e.g.) [creaky voice, in = 4, sp = 6] was intended to represent *k'a'an* in order to test the interim grammar.

Based on the discussion above and in §5.2.3, it is clear that perception learning will teach the child the relations among the members of each set of constraints that refer to a certain [x], but the child will not learn the relations among the members of each set of constraints that refer to a certain /y/. If it is thus the case that the mean ranking values of cue constraints are learned through perception learning alone, this predicts that learners only need to learn the relations among each [x] set and that the correct production grammar will simply fall out of a grammar that is developed through perception learning. Before testing this idea with the data from Yucatec Maya, we will discuss the limitations of this learning strategy and some possible types of language data that would be problematic for it.

Referring back to the tableau in (5.4), we see that, for each constraint that is demoted, some other constraint (which refers to the same [x] but a different /y/) is promoted. For example, */gl/,[in = 4] is demoted while */hi/,[in = 4] is promoted. This will always be the case when perception learning is done with a grammar that uses cue constraints to penalize the pairing of each possible phonetic value with each possible surface form and there are only two surface forms; the demotion of */y_1/,[x] will always be accompanied by a promotion (of equal value) of */y_2/,[x] (and vice versa). This means that the ranking of the cue constraints that regulate one surface form will mirror the ranking of the cue constraints that regulate the other surface form. Thus, if our initial state is one in which all constraints have the mean ranking value of 100, perception learning will lead to a grammar where the distance between the mean ranking of */y_1/,[x] and 100 is equal (though opposite in direction) to the distance between the mean ranking value of */y_2/,[x] and 100.

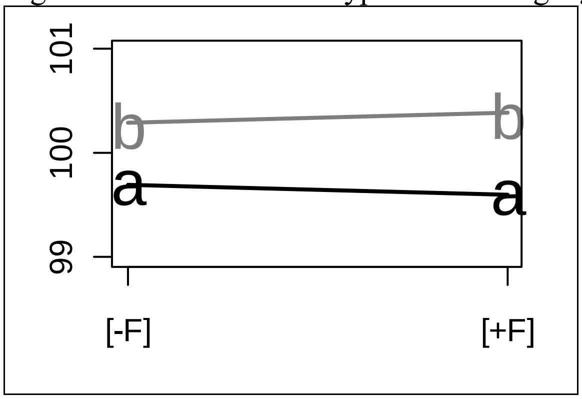
We can imagine a type of language data for which this would not accurately reflect the facts about both production and perception. Suppose some language has the two surface forms /a/ and /b/, that /a/ is more frequent than /b/, and that both /a/ and /b/ are produced as [-F] 90% of the time (and as [+F] 10% of the time). This data is summarized in (5.5)

(5.5) hypothetical language

/SF/	[PF]	quantity
/a/	[-F]	70
/a/	[+F]	7
/b/	[-F]	50
/b/	[+F]	5

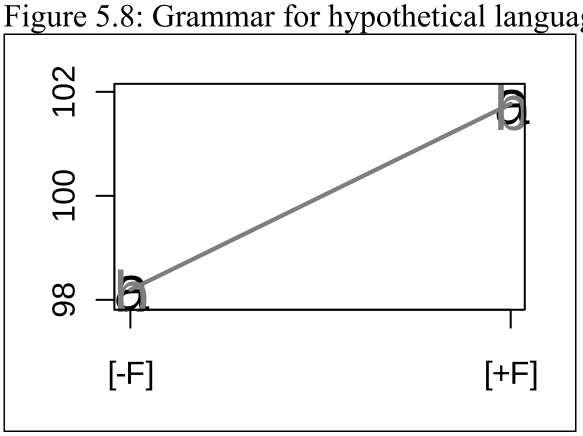
Given such a language, we may expect that, in perception, the listener would be more likely to map both [-F] and [+F] onto /a/ and that, in production, the speaker would be more likely to map both /a/ and /b/ onto [-F]. However, such a grammar cannot be learned through perception learning alone. If we use this language data to run a learning simulation with PRAAT where the learner is trained on perception tableaus alone, we get the grammar shown in Fig. 5.7. This grammar says that, in perception, both [-F] and [+F] are more likely to be mapped onto /a/ (as desired), but that, in production, both [-F] and [+F] are almost equally likely to be produced for either /a/ or /b/ (with [+F] being slightly more likely for /a/ and [-F] being slightly more likely for /b/). This production grammar does not capture the facts about the language as shown in (5.5), specifically in that [-F] should be far more likely to be produced for either surface form. In this case, perception learning has not led to an accurate production grammar. The reasons that perception learning fails will be discussed in more detail at the end of this section.

Figure 5.7: Grammar for hypothetical language developed through perception learning



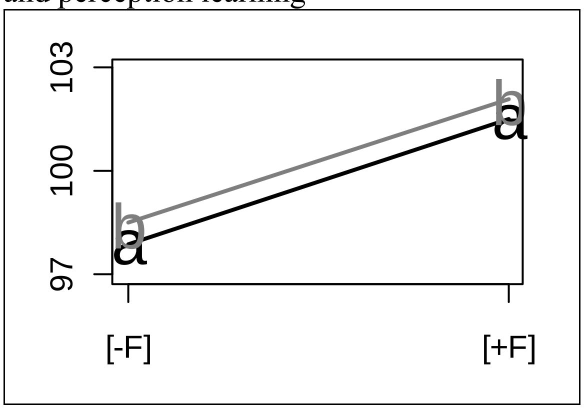
It is also clear that, with our hypothetical language, production learning alone does not lead to an accurate perception grammar. Using the same data, a learning simulation was run in which the learner was trained on production tableaus alone, and the resulting grammar is shown in Fig. 5.8. In this grammar the mean ranking values of */a/, [-F] and */b/[-F] are almost identical (as are */a/,[+F] and */b/,[+F]) and hence the two lines of the graph are right on top of each other. This time, the production grammar is accurate: both /a/ and /b/ are more likely to be produced as [-F]. However, the perception grammar now predicts that either [-F] or [+F] is equally likely to be mapped onto /a/ or /b/, which does not accurately reflect the data in (5.5)

Figure 5.8: Grammar for hypothetical language developed through production learning



The only way for the learner to reach the desired grammar is for that learner to be trained on both production and perception tableaus. In order to do this one more learning simulation was run in which the learner tests both production and perception tableaus at each learning step. The grammar that results from this learning simulation is shown in Fig. 5.9. This perception grammar predicts for both [-F] and [+F] to be more likely to be mapped onto /a/, and this production grammar predicts that both /a/ and /b/ to be more likely to be mapped onto [-F].

Figure 5.9: Grammar for hypothetical language developed through simultaneous production and perception learning



A learner could not learn the hypothetical language in (5.5) through perception learning alone because this learning strategy cannot lead to a grammar where both */a/,[-F] and */b/,[-F] are ranked below the initial ranking of 100 and where both */a/,[+F] and */b/,[+F] are ranked above the initial ranking of 100. With perception learning, every time */a/[-F] is demoted, */b/,[-F] is promoted. However, because [-F] is good phonetic form for both /a/ and /b/, the correct adult grammar must have both */a/[-F] and */b/,[-F] with relatively low mean ranking values, which cannot be achieved through perception learning alone.

This discussion has pointed to some of the limitations of learning the ranking values of cue constraints through perception (or production) alone. An issue that we have not yet addressed is the ranking of articulatory constraints. While perception learning can be used for the rankings of cue constraints, it cannot have an impact on the rankings of articulatory constraints. Boersma and Hamann (2008) use a model with fixed, universal (i.e. not susceptible to language-specific learning) rankings for articulatory constraints (based on Kirchner 1998), where the rankings are determined by how effortful the articulation is. In a footnote they say that some language-specific learning may affect the rankings of articulatory constraints insofar as 'practiced' articulations become less effortful and hence are not penalized as severely (following Boersma 1998).

I have no argument against the theory behind this proposal, but it is hard to implement in our current model. If articulatory constraints have these fixed rankings, how do we determine what they are? I do not know of any research in the context of the Bidirectional StOT that has set out to answer this question, and it is certainly beyond the scope of this dissertation do so.[64] This is why I proposed the scalar articulatory constraints in (5.3). Many of the same insights are shared by de Lacy's (2002) proposals about markedness and Kirchner's (1998) proposals about the role of effort, and for this reason I feel that using freely rankable scalar constraints is an appropriate substitution. This means that the learner will have to learn the rankings of the articulatory constraints. I will try two different strategies for learning the rankings of articulatory constraints, one in which production learning is used to learn only the rankings of articulatory constraints (while cue constraints retain the ranking values developed through perception learning) and one in which production learning is used to learn the rankings of both articulatory and cue constraints. We will see that the second strategy works best, but that both strategies lead to rankings of

[64] In Boersma and Hamann (2008) and Boersma (2006), where articulatory constraints are used for analysis, only one phonetic dimension is evaluated and the data to be accounted for is simulated to reflect a normal distribution, which makes it easier to propose values for articulatory constraints that work within the context of the particular analysis. In my analysis, where multiple phonetic dimensions are in play and a set of articulatory constraints refers to two dimensions at once, it seems inappropriate to invent ranking values for the articulatory constraints instead of using language-specific learning to adjust the ranking values.

articulatory constraints that would be problematic for a system in which no language-specific learning is used with articulatory constraints (see §5.7).

In the following sections we will see that there are patterns in the Yucatec Maya data that resemble the hypothetical language in (5.5) and thus that are problematic for a learning strategy that trains the learner with only perception tableaus. In order to better understand the difference between perception learning alone and combined production and perception learning, we will develop two different grammars in the rest of this chapter. The first grammar will make use of perception learning (as proposed by Boersma (2006)) for ranking the cue constraints and production learning for ranking the articulatory constraints. The second grammar will make of use the same learning strategy that was used to develop the grammar in Fig. 5.9 – simultaneous production and perception learning (which affects the rankings of both cue and articulatory constraints). We will see that simultaneous production and perception learning leads to a more accurate production grammar.

5.4 Yucatec Maya Grammar 1: Perception Learning

In this section we develop a grammar by using perception learning (Boersma 2006, see last section) to rank the cue constraints. I first discuss the details of the learning strategy in §5.4.1 and then present the constraint rankings that result in §5.4.2. We will then assess the success of the adult grammar that results from the learning simulations by comparing the predicted output distributions of both the production and perception grammars to the data collected from actual speakers and listeners of Yucatec Maya in §5.4.3. We will have to be cautious in our assessment of the production grammar, as perception learning cannot have any influence over the ranking of articulatory constraints. Articulatory constraints are added to the model in §5.5.

5.4.1 The Learning Strategy

In order to develop this grammar, the strategy of learning through perception tableaus (within the context of the GLA) is used to rank the cue constraints that were presented in Table 5.6. The data that the learner is trained on comes from the production studies, as discussed in §5.1. Each token from the production studies indicates both a phonetic form (as defined by the dimensions of glottalization type, initial pitch, and pitch span) and a surface form (the phonological category of /gl/ or /hi/). In this way, the learner hears some [x] and knows that it should map onto some specific /y/, and the learner can thus test their current perception grammar to see if it makes the correct prediction.

Phonetic forms are defined by three different phonetic dimensions (in our Yucatec Maya grammar). If we consider these dimensions all at once (i.e. by evaluating a full form like [creaky voice, in = 4, sp = 6]) the GLA will perform different ranking adjustments than if we consider each dimension individually (i.e. by evaluating the form [creaky voice] separately from the form [in=4], etc.). The fact that this is so can be illustrated with the same tableau that was used to illustrate perception learning above in (5.4), repeated below in (5.6). In this tableau, /gl/ is what the speaker said (with the form [creaky voice, in = 4, sp = 6]), but the current perception grammar maps that input onto /hi/. The constraint */gl/, [in=4] assigns the fatal violation, and hence the ranking values of all constraints shown are adjusted in accord with the GLA.

(5.6) adjustment of ranking values when one input specifies each phonetic dimension

	→	←	←	←	→	→
[creaky voice, in =4, sp = 6]	*/gl/, [in=4]	*/hi/, [cr]	*/hi/, [in=4]	*/hi/, [sp=6]	*/gl/, [sp=6]	*/gl/, [cr]
☞ /hi/		*	*	*		
☜ /gl/	*!				*	*

However, the phonetic values of [creaky voice] and [sp=6] are preferably mapped onto GLOTTALIZED vowels according to the interim grammar. If we consider each phonetic dimension separately, as shown in (5.7), only the ranking values for the relevant constraints that reference initial pitch are adjusted. Thus we see that the two different strategies could result in different mean ranking values for the cue constraints. Which strategy is the best? On one hand, it seems that the learner should look at the whole phonetic form at once, and this would be necessary if there are single cue constraints that reference multiple phonetic dimensions. On the other hand, it seems undesirable for a mismatch between initial pitch and surface form to affect the ranking of constraints that reference glottalization and pitch span.

(5.7) adjustment of ranking values when three different inputs specify only one phonetic dimension each

a. initial pitch

	→	←
[in=4]	*/gl/, [in=4]	*/hi/, [in=4]
☞ /hi/		*
✋ /gl/	*!	

b. glottalization

[creaky voice]	*/hi/, [cr]	*/gl/, [cr]
/hi/	*!	
✋,☞ /gl/		*

c. pitch span

[sp=6]	*/hi/, [sp=6]	*/gl/, [sp=6]
/hi/	*!	
✋,☞ /gl/		*

We thus have to determine which approach the GLA learner should take: evaluate all phonetic dimensions simultaneously as in (5.6) or evaluate each dimension individually as in (5.7).

The nature of the data I am using may have complications for the method where all phonetic dimensions are evaluated at once. Because my data set includes 685 tokens of natural speech, there are of course legitimate combinations of specific values for the phonetic dimensions that are missing. For example, no one produced a phonetic form of [weak glottalization, in = -2, sp = 2], and so there is no data to train the learner on how to perceive this form. The learner will of course learn the ranking values of all the relevant cue constraints, and hence the grammar would be able to evaluate this phonetic form. However, the fact remains that if one such token had been produced, the learner would be trained on that token and this training could affect the ranking values of all the cue constraints involved. Thus, the fact that my data set contains accidental gaps could have consequences for a learning simulation where all phonetic dimensions are evaluated simultaneously. If we look at each phonetic dimension individually, there are no accidental gaps; every category of glottalization type, initial pitch, and pitch span was repeatedly produced.

In order to determine how the adult grammar that results from GLA learning is affected by the two different approaches, a learning simulation was run with each method. The two resulting grammars were quite different. The one that resulted from evaluating each dimension separately captures the main generalizations that we have made about the data set and is presented in this section. The one that resulted from the evaluation of every dimension simultaneously did not, for the most part, describe the patterns of the data. I will present this grammar and discuss the differences between the two in §6.1.3. Because of the accidental gaps in the data when all dimensions are evaluated simultaneously, I cannot be sure if this learning strategy is flawed or just ill-suited to the data set.

5.4.2 Results of the Learning Simulation

The current section presents the results of a learning simulation where perception learning is used to learn the rankings of cue constraints (as described in the preceding section). At each learning step, the learner takes a learning datum, which is some phonetic form (defined by only one phonetic dimension) that a speaker said for some surface form. The learner checks to see if the current grammar predicts for that phonetic form to be mapped onto the surface form that the speaker said. If the grammar makes an incorrect prediction, a ranking adjustment occurs. Because the input indicates a value for only one phonetic dimension, only the constraints that regulate this dimension can be adjusted in any one step.

The desired frequency of each surface form – phonetic form pairing is determined by the data set of Yucatec Maya speech. All PRAAT defaults are used to run this learning simulation

The mean ranking values that result from this learning simulation are provided in Table 5.8 and displayed in Fig. 5.10.

Figure 5.10: Mean ranking values of cue constraints (developed through perception learning)

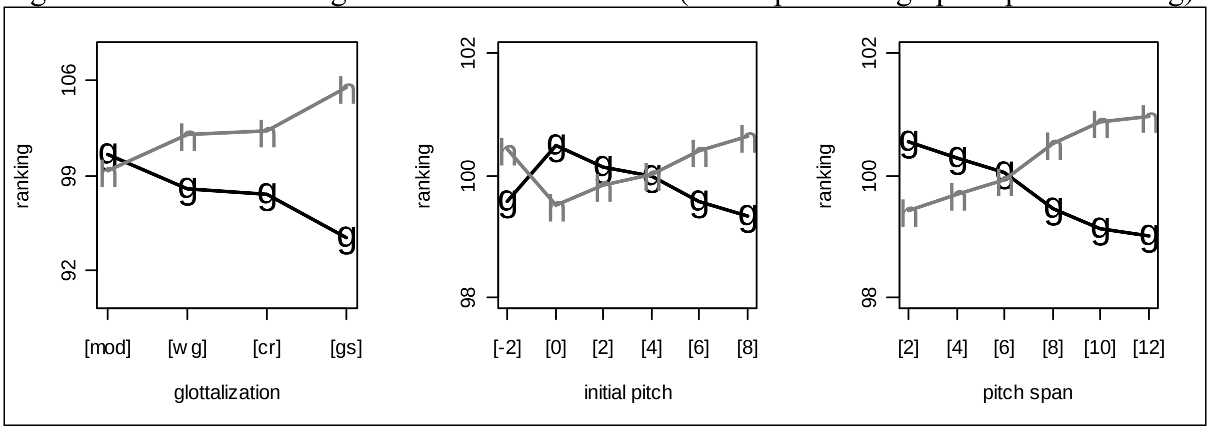


Table 5.8: Mean ranking values of cue constraints (developed through perception learning)

	constraint	ranking value	constraint	ranking value
glottalization	*/gl/,[mod]	100.59	*/hi/,[mod]	99.40
	*/gl/,[wg]	97.99	*/hi/,[wg]	102.00
	*/gl/,[cr]	97.65	*/hi/,[cr]	102.34
	*/gl/,[gs]	94.47	*/hi/,[gs]	105.52
initial pitch	*/gl/,[in = -2]	99.57	*/hi/,[in = -2]	100.42
	*/gl/,[in = 0]	100.48	*/hi/,[in = 0]	99.51
	*/gl/,[in = 2]	100.15	*/hi/,[in = 2]	99.84
	*/gl/,[in = 4]	99.98	*/hi/,[in = 4]	100.01
	*/gl/,[in = 6]	99.58	*/hi/,[in = 6]	100.41
	*/gl/,[in = 8]	99.34	*/hi/,[in = 8]	100.65
pitch span	*/gl/,[sp = 2]	100.56	*/hi/,[sp = 2]	99.43
	*/gl/,[sp = 4]	100.29	*/hi/,[sp = 4]	99.70
	*/gl/,[sp = 6]	100.05	*/hi/,[sp = 6]	99.94
	*/gl/,[sp = 8]	99.46	*/hi/,[sp = 8]	100.53
	*/gl/,[sp = 10]	99.12	*/hi/,[sp = 10]	100.87
	*/gl/,[sp = 12]	99.03	*/hi/,[sp = 12]	100.96

5.4.3 Assessment of the Grammar

In evaluating the success of this learning simulation, we wish to consider the predictions that are made with respect to both the production and perception grammars.

Because this learning simulation involved only training through perception tableaus, we will consider the perception grammar first. If we look at Fig. 5.10, we can generalize about the perception grammar by considering the relation between two constraints in a given column. For example, this grammar demonstrates a strong preference for mapping a [glottal stop] onto a GLOTTALIZED vowel (because */gl/,[gs] has a much lower mean ranking value than */hi/,[gs]). If we look at each column, we see that those phonetic values that occur more often with GLOTTALIZED vowels (glottalization, higher initial pitch, larger pitch span) will trigger the perception of a GLOTTALIZED vowel and those phonetic values that occur more often with HIGH TONE vowels will trigger the perception of a HIGH TONE vowel. Thus, to the human linguist, the generalizations about the perception grammar that we can make based on Fig. 5.10 seem to match the generalizations that we made about the data set.

In Table 5.9 we see how often this grammar predicts phonetic forms (as defined by all three phonetic dimensions) to be mapped onto GLOTTALIZED vowels. Because a HIGH TONE vowel is the only other option, if a GLOTTALIZED vowel is not deemed optimal by the grammar, a HIGH TONE vowel is. Hence, the table shows that [modal voice, in = -2, sp = 2] is mapped onto a GLOTTALIZED vowel 41% of the time, and we can deduce that this phonetic form is mapped onto a HIGH TONE vowel 59% of the time.

Table 5.9: Percentage of times the perception grammar predicts a GLOTTALIZED vowel to be the optimal output for a given phonetic input

	pitch span	initial pitch					
		-2	0	2	4	6	8
modal	2	41	30	34	36	41	44
	4	45	33	37	39	44	47
	6	47	36	40	42	47	50
	8	55	43	47	49	54	57
	10	59	48	51	54	59	61
	12	59	49	53	55	59	62
weak glottalization	2	71	63	66	68	71	73
	4	74	65	69	71	74	76
	6	77	68	71	73	77	79
	8	82	73	77	79	82	84
	10	85	76	80	81	85	86
	12	85	77	80	82	85	87
creaky	2	75	67	70	71	75	76
	4	77	69	72	74	77	79
	6	80	71	75	76	80	81
	8	85	76	80	81	85	86
	10	87	79	82	84	87	88
	12	87	79	83	84	87	89
glottal stop	2	95	94	94	95	96	96
	4	96	94	95	95	96	96
	6	97	95	96	96	97	97
	8	98	96	97	97	98	98
	10	98	96	97	97	98	98
	12	98	96	97	97	98	98

Do the predictions match the behavior of actual listeners as documented by task 1 of the perception study? The results (for Santa Elena only) are reproduced below as Table 5.10. Listeners are more likely to select a GLOTTALIZED vowel if the stimulus has more glottalization, and/or a larger pitch span. For initial pitch, extremely low values are associated with more GLOTTALIZED vowel responses, but otherwise higher initial pitch values trigger more GLOTTALIZED vowel responses. This is exactly the pattern that is predicted by our perception grammar.[65] There are two minor areas where Table 5.9 and Table 5.10

[65] Unfortunately, because the stimuli for task 1 only included 24 tokens of each surface form, it would not be meaningful to use those stimuli and the participants' responses to create a full table of observed responses for each combination of glottalization type, initial pitch value, and pitch span as in Table 5.9. We should also be cautious about interpreting the exact numbers in the cells of Table 5.10. Because we only have enough data to look at each phonetic dimension individually, we cannot measure the interactions of the phonetic dimensions,

diverge: real listeners selected more GLOTTALIZED vowels for the weak glottalization category than for the creaky voice category and real listeners were less likely to select a GLOTTALIZED vowel for the initial pitch category of [8].[66]

Table 5.10: Review of results for task 1 (Santa Elena only)
Percentage of times a GLOTTALIZED vowel was chosen on the basis of stimulus glottalization type, initial pitch, and pitch span. (Repeated from Table 5.6)

glottalization type					
ng	wg	cr	gs		
39	68	55	93		
initial pitch category					
-2	0	2	4	6	8
52	44	38	59	79	50
pitch span category					
2	4	6	8	10	12
33	41	59	63	71	84

It thus seems that perception learning has resulted in the development of an accurate perception grammar. When we evaluate the success of the production grammar, things are not so straightforward. First, I will discuss the patterns that are predicted to occur based on the graphical representation of the constraint rakings in Fig. 5.10. Then I will present a more formal analysis where we will use the χ^2 statistic to measure the degree of divergence between the distribution of phonetic forms in the production data and the distribution that is predicted to occur based on our production grammar.

Before considering the rankings of the cue constraints as shown in Fig. 5.10, we need to remember that articulatory constraints have not yet been introduced into the model. If we look at the constraints that regulate glottalization type, the production grammar defined by

---

and so the numbers for any one phonetic dimension can only be used to define behavior trends. For example, there is a clear trend that more glottalization triggers more GLOTTALIZED vowel responses, but we cannot conclude that real listeners will map a glottal stop onto a GLOTTALIZED vowel 93% of the time because we cannot be certain how real listeners handle the interaction of glottalization type and initial pitch.

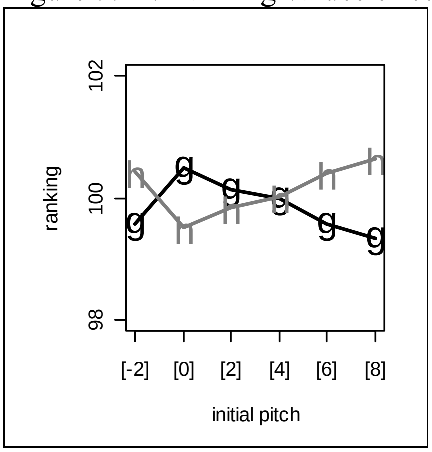
[66] Recall that the fact that Santa Elena listeners selected a GLOTTALIZED vowel when the initial pitch was [8] is anomalous; listeners from Mérida selected GLOTTALIZED vowels 100% of the time when the initial pitch value was [8].

cue constraints predicts that a GLOTTALIZED vowel will most often be mapped onto a production with a [glottal stop]. We of course know that this is not true in Yucatec Maya; a full glottal stop is rarely produced. This problem can be remedied with the articulatory constraint *[glottal stop]. As long as this articulatory constraint has a higher ranking value than at least one of the constraints */gl/, [cr]; */gl/,[wg]; or */gl/,[mod], an input of a GLOTTALIZED vowel is more likely to be mapped onto another glottalization type than to be mapped onto a production with a glottal stop. In general, a production grammar defined by cue constraints alone is not likely to be very successful because the articulatory constraints do important work in the grammar.

Thus, the truly problematic parts of the production grammar are those parts that cannot be fixed by articulatory constraints. There are two main areas where an incorrect prediction is made for the production grammar which I believe cannot be repaired with articulatory constraints.

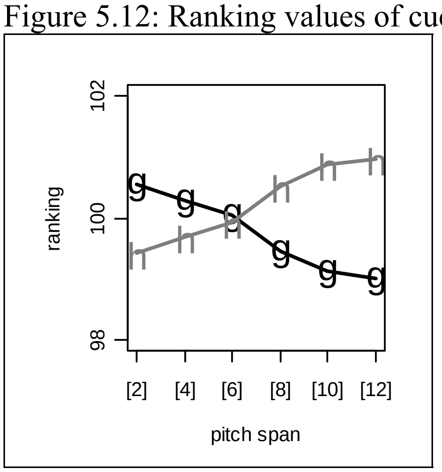
First, if we look at the constraints that regulate initial pitch (repeated below in Fig. 5.11), we see that [0] is predicted to be the preferred value for initial pitch for an input of a HIGH TONE vowel. Though this is a common initial pitch value for HIGH TONE vowels, this is not the value that occurs most often in production – both [2] and [4] are more prevalent in the data. Furthermore, an initial pitch of [0], which is right at a speaker's baseline, should not be very effortful, so we cannot rely on articulatory constraints to rule it out. In fact, both [2] and [4] should be more effortful, and so the presence of articulatory constraints would make their productions even less likely. Another potential problem is how much the initial pitch values of [0] and [2] are disprefered for GLOTTALIZED vowels, even though these values are commonly produced. In this case, articulatory constraints could rule out the higher pitch values, making [0] and [2] more likely to be produced, but these articulatory constraints would also rule out these pitch values for HIGH TONE vowels. This would make [0] even more likely for HIGH TONE vowels, which was the original problem that was indicated.

Figure 5.11: Ranking values of cue constraints that regulate initial pitch



If we look at the ranking values of the constraints that regulate pitch span (repeated below in Fig. 5.12), we see a prediction that the preferred pitch span for HIGH TONE vowels is [2]. Again, though, this is a fairly unmarked value and it is not the most frequent production. In the data set, HIGH TONE vowels are most commonly produced with a pitch span of [4], which is more marked value.

Figure 5.12: Ranking values of cue constraints that regulate pitch span



In order to present a formal analysis of the accuracy of the learned production grammar, Table 5.11 presents the values used to calculate a χ^2 statistic as a measure of the degree of how greatly the observed frequencies of input-output pairings (from the production studies) diverge from the expected frequencies (as predicted by the production grammar developed with the GLA using perception learning). In this table, the "obs. freq." column gives the number of times (out of 685 total tokens) that a given input was paired with a given output. Because these values represent the exact numbers of tokens that were actually produced, they are always whole numbers. The "exp. freq." column gives the number of times (out of 685) that a given output is predicted to occur with a given input. These numbers are calculated with PRAAT, which determines the number of times out of 100,000 that a given output is deemed optimal by the grammar for a given input. In order to calculate the χ^2 statistic, the expected frequency value out of 100,000 trials as obtained from PRAAT was transformed into expected frequency out of 685. For this reason, the numbers in the "exp. freq." column are not whole numbers. As presented, they are rounded to the nearest hundredth, but the unrounded number was used to calculate the χ^2 statistic. The numbers in the χ^2 column are the terms of the χ^2 statistic, as calculated with the equation (obs. value – exp. value)^2 / obs. value. If we sum all of the values in the χ^2 column, we get the χ^2 statistic: 70,183.88.

Because the expected values are essentially fixed (n = 100,000), we cannot use this statistic for hypothesis testing, but it will be compared to the χ^2 statistic obtained with different grammars in the rest of this chapter as a way to compare how successful the various grammars are at accounting for the production data. The smaller the number, the less that observed and expected frequencies diverge. It is clear that this number is not small and hence that the expected output distributions diverge greatly from the observed output distributions.

If we look for the cells with greatest degree of divergence, we see that, as expected, this grammar predicted too many productions of GLOTTALIZED vowels with a glottal stop and a too few productions with modal voice. As discussed above, this problem could be remedied with articulatory constraints. However, also as discussed above, there are problematic areas for which articulatory constraints are not an ideal solution. For example, too many HIGH TONE vowels are predicted to be produced with an initial pitch [0] and not enough are predicted to be produced with an initial pitch of [2].

Table 5.11: Comparison of observed and expected frequencies of input and output pairings (perception learning)
$\chi^2 = 70,183.88$

a. surface form = /gl/

		mod			wg			cr			gs		
init pitch	pitch span	obs. freq.	exp. freq.	χ^2	obs. freq.	exp. freq.	χ^2	obs. freq.	exp. freq.	χ^2	obs. freq.	exp. freq.	χ^2
-2	2	4	0.11	138.10	0	1.01	1.01	2	1.26	0.43	0	8.95	8.95
	4	4	0.08	186.73	0	1.09	1.09	2	1.71	0.05	0	11.05	11.05
	6	1	0.14	5.44	0	1.40	1.40	1	1.70	0.29	0	13.56	13.56
	8	2	0.20	16.33	1	1.99	0.49	2	2.77	0.22	1	20.74	18.79
	10	0	0.21	0.21	1	2.86	1.21	0	3.82	3.82	0	26.92	26.92
	12	5	0.29	77.18	0	2.82	2.82	3	3.88	0.20	0	28.35	28.35
0	2	16	0.03	7442.49	1	0.49	0.54	1	0.64	0.20	0	4.30	4.30
	4	5	0.05	446.26	1	0.56	0.34	2	0.69	2.52	0	5.51	5.51
	6	0	0.07	0.07	1	0.70	0.13	1	0.94	0.00	1	6.74	4.89
	8	3	0.10	87.94	2	1.05	0.86	3	1.51	1.46	0	10.25	10.25
	10	0	0.08	0.08	0	1.11	1.11	1	1.80	0.36	0	13.72	13.72
	12	0	0.14	0.14	1	1.41	0.12	3	1.93	0.59	0	13.77	13.77
2	2	20	0.05	8302.07	1	0.64	0.20	3	0.73	7.12	1	5.39	3.58
	4	34	0.03	33683.86	4	0.80	12.77	2	0.83	1.65	0	6.77	6.77
	6	8	0.10	651.46	7	0.83	45.95	2	1.17	0.59	0	8.22	8.22
	8	6	0.11	316.58	1	1.40	0.11	9	1.99	24.63	0	13.25	13.25
	10	1	0.18	3.59	0	1.86	1.86	2	2.49	0.10	0	17.32	17.32
	12	1	0.14	5.10	1	2.05	0.54	13	2.40	46.89	2	17.78	14.00
4	2	2	0.06	60.94	0	0.60	0.60	1	1.07	0.00	0	6.40	6.40
	4	24	0.04	13966.64	4	0.79	12.93	3	1.14	3.01	0	8.01	8.01
	6	14	0.12	1561.74	5	0.99	16.16	6	1.56	12.61	0	10.02	10.02
	8	5	0.08	321.86	5	1.52	7.96	3	2.23	0.27	1	14.77	12.84
	10	0	0.14	0.14	1	1.96	0.47	1	2.65	1.03	0	20.05	20.05
	12	0	0.17	0.17	2	2.08	0.00	11	2.86	23.22	0	21.12	21.12
6	2	1	0.10	8.52	0	0.87	0.87	0	1.32	1.32	0	8.47	8.47
	4	5	0.10	250.78	1	1.13	0.02	0	1.45	1.45	0	10.58	10.58
	6	10	0.09	1103.05	3	1.37	1.94	4	1.80	2.68	0	13.66	13.66
	8	5	0.21	107.94	1	2.18	0.64	1	2.91	1.25	0	20.45	20.45
	10	2	0.25	12.47	1	2.72	1.09	1	3.75	2.01	0	25.43	25.43
	12	1	0.25	2.20	1	2.89	1.24	5	3.56	0.59	0	28.64	28.64
8	2	1	0.09	9.32	0	1.12	1.12	0	1.29	1.29	0	10.10	10.10
	4	2	0.14	25.33	1	1.20	0.03	0	1.86	1.86	0	12.46	12.46
	6	0	0.10	0.10	3	1.62	1.18	0	2.21	2.21	0	15.73	15.73
	8	5	0.24	94.52	2	2.44	0.08	0	3.40	3.40	0	24.84	24.84
	10	3	0.25	29.76	0	3.22	3.22	1	4.51	2.73	0	30.85	30.85
	12	0	0.27	0.27	0	3.50	3.50	0	4.73	4.73	0	33.19	33.19

Table 5.11, continued

b. surface form = /hi/

		mod			wg			cr			gs		
init pitch	pitch span	obs. freq.	exp. freq.	χ^2	obs. freq.	exp. freq.	χ^2	obs. freq.	exp. freq.	χ^2	obs. freq.	exp. freq.	χ^2
-2	2	2	18.58	14.79	0	3.75	3.75	1	2.90	1.24	0	0.16	0.16
	4	8	14.80	3.12	0	3.10	3.10	0	2.16	2.16	0	0.08	0.08
	6	2	12.47	8.79	0	2.51	2.51	0	1.78	1.78	0	0.12	0.12
	8	1	7.60	5.73	0	1.50	1.50	0	1.21	1.21	0	0.03	0.03
	10	1	5.85	4.02	0	1.23	1.23	0	1.01	1.01	0	0.03	0.03
	12	2	5.51	2.24	0	1.10	1.10	0	0.82	0.82	0	0.05	0.05
0	2	30	35.45	0.84	0	7.08	7.08	0	5.40	5.40	0	0.18	0.18
	4	30	28.75	0.05	0	5.45	5.45	0	4.51	4.51	0	0.22	0.22
	6	5	25.14	16.13	0	5.06	5.06	0	4.03	4.03	0	0.17	0.17
	8	6	15.45	5.78	0	3.21	3.21	0	2.46	2.46	0	0.14	0.14
	10	0	11.73	11.73	0	2.27	2.27	0	1.89	1.89	0	0.08	0.08
	12	3	10.99	5.81	0	2.00	2.00	0	1.69	1.69	0	0.05	0.05
2	2	51	27.42	20.28	1	5.66	3.83	0	4.27	4.27	0	0.24	0.24
	4	56	22.75	48.60	0	4.71	4.71	0	3.66	3.66	0	0.23	0.23
	6	17	19.73	0.38	0	4.19	4.19	0	3.06	3.06	0	0.16	0.16
	8	6	12.36	3.28	0	2.65	2.65	0	1.85	1.85	0	0.12	0.12
	10	0	9.22	9.22	0	1.95	1.95	0	1.49	1.49	0	0.06	0.06
	12	3	8.38	3.45	1	1.73	0.31	1	1.38	0.11	0	0.03	0.03
4	2	13	24.04	5.07	0	4.64	4.64	0	3.86	3.86	0	0.18	0.18
	4	28	21.17	2.21	0	4.08	4.08	0	3.42	3.42	0	0.16	0.16
	6	33	17.52	13.69	0	3.45	3.45	0	2.73	2.73	0	0.09	0.09
	8	9	11.21	0.44	0	2.23	2.23	0	1.74	1.74	0	0.08	0.08
	10	1	8.63	6.75	1	1.64	0.25	0	1.40	1.40	0	0.03	0.03
	12	0	7.75	7.75	1	1.62	0.24	2	1.32	0.35	0	0.03	0.03
6	2	0	18.08	18.08	0	3.73	3.73	0	2.64	2.64	0	0.19	0.19
	4	6	15.41	5.74	0	2.82	2.82	0	2.27	2.27	0	0.13	0.13
	6	10	12.67	0.56	0	2.48	2.48	0	1.97	1.97	0	0.10	0.10
	8	6	8.38	0.67	0	1.68	1.68	0	1.14	1.14	0	0.05	0.05
	10	2	6.18	2.83	0	1.17	1.17	0	0.81	0.81	0	0.02	0.02
	12	0	5.44	5.44	0	1.17	1.17	1	0.90	0.01	0	0.04	0.04
8	2	0	14.55	14.55	0	2.91	2.91	0	2.40	2.40	0	0.09	0.09
	4	0	12.36	12.36	0	2.50	2.50	0	2.00	2.00	0	0.10	0.10
	6	3	10.47	5.33	0	1.93	1.93	0	1.53	1.53	0	0.09	0.09
	8	3	6.72	2.06	0	1.32	1.32	0	0.97	0.97	0	0.05	0.05
	10	0	5.06	5.06	0	1.16	1.16	0	0.79	0.79	0	0.06	0.06
	12	2	4.66	1.52	0	0.90	0.90	0	0.84	0.84	0	0.03	0.03

5.5 Adjusting Yucatec Maya Grammar 1: Learning the Rankings of Articulatory Constraints

Before we can truly evaluate the success of perception learning in developing accurate rankings of the cue constraints, we must add articulatory constraints to our model. In the previous section, we used perception learning to learn the rankings of cue constraints. We saw that the production grammar that is defined by these cue constraints was not very accurate, but it was suggested that many of the inaccuracies could be remedied with the articulatory constraints. For example, too many GLOTTALIZED vowels were predicted to be produced with a [glottal stop], but if we include *[glottal stop] in our grammar, many of these productions could incur a fatal violation of this constraint. It is thus important to add articulatory constraints to the model before reaching any conclusions about the quality of the production grammar that was developed through perception learning.

In this section, I add articulatory constraints to our model. I use a learning strategy where the learner is trained on production tableaus only, where the articulatory constraints begin with a mean ranking value of 100, and where the cue constraints have a plasticity of 0. This means that, no matter what errors the interim production grammar makes, the ranking values of the cue constraints will never be altered; errors can only affect the ranking values of articulatory constraints. In this way, the ranking values of cue constraints are solely learned through perception learning (as is claimed by Boersma 2006), but then language-specific production learning is used to rank the articulatory constraints (*contra* Boersma 2006 and Boersma and Hamann 2008). The mean ranking values of the articulatory constraints that result from this learning simulation are given in Table 5.12.

Table 5.12: Mean ranking values of articulatory constraints (perception learning of cue constraints, production learning of articulatory constraints)

*[adducted vocal folds]	*[wg]	102.87
	*[cr]	102.04
	*[gs]	105.15
*[creaky voice, small pitch span]	*[cr, sp<6]	99.49
	*[cr, sp<4]	99.67
	*[cr, sp<2]	98.97
*[large pitch span]	*[sp>4]	-5607.68
	*[sp>6]	97.99
	*[sp>8]	98.89
	*[sp>10]	98.74
	*[sp>12]	-3467.26

Because the mean ranking values of cue constraints have not been altered by this learning simulation, there has not been any change to the perception grammar. We already know that this perception grammar is highly accurate, so we only need to evaluate the success of the new production grammar. In order to do this, PRAAT is used to predict how many times (out of 100,000) each surface form is mapped onto each phonetic form, and the χ^2 statistic is calculated as before. The result is $\chi^2 = 825.74$, as shown in Table 5.13. As expected, the addition of articulatory constraints has led to a more accurate production grammar (826 < 70,184). One area of the grammar that has drastically improved is that many fewer productions of a full glottal stop are predicted to occur, which accurately reflects the data.

However, there are still some problematic areas for the new grammar. As discussed above, the ranking of cue constraints (as learned through perception learning) suggests that an initial pitch of [0] should be highly dispreferred for GLOTTALIZED vowels, even though this value commonly occurs in productions of GLOTTALIZED vowels. Because there are no articulatory constraints in our model that penalize initial pitch values, this new grammar still predicts too few GLOTTALIZED vowels to be produced with [in = 0]. Another problem that was not solved with the addition of articulatory constraints is that too few GLOTTALIZED vowels are predicted to be produced with a pitch span of [12], e.g. the token [cr, in = 2, sp = 12] occurs 13 times but is only predicted to occur 3 times.

The problems with the production grammar that cannot be fixed with articulatory constraints are subtle, but they are there. In the next section we will seek to improve the grammar by allowing production learning to also alter the rankings of cue constraints. This new learning step will lead to an improved production grammar and thus show that the learner cannot train on perception tableaus alone and learn the appropriate relations among constraints of the /gl/ set or the /hi/ set.

Table 5.13: Comparison of observed and expected frequencies of input and output pairings (perception learning + production learning of articulatory constraints)
$\chi^2 = 825.74$

a. surface form = /gl/

		mod			wg			cr			gs		
init pitch	pitch span	obs. freq.	exp. freq.	χ^2	obs. freq.	exp. freq.	χ^2	obs. freq.	exp. freq.	χ^2	obs. freq.	exp. freq.	χ^2
-2	2	4	20.35	13.14	0	1.00	1.00	2	2.06	0.00	0	0.59	0.59
	4	4	24.71	17.36	0	2.02	2.02	2	3.69	0.77	0	0.70	0.70
	6	1	18.84	16.89	0	3.43	3.43	1	5.90	4.07	0	0.59	0.59
	8	2	10.87	7.24	1	5.29	3.48	2	9.36	5.78	1	0.24	2.41
	10	0	5.11	5.11	1	2.63	1.01	0	4.41	4.41	0	0.18	0.18
	12	5	5.60	0.06	0	2.77	2.77	3	4.71	0.62	0	0.12	0.12
0	2	16	10.23	3.25	1	0.53	0.42	1	0.98	0.00	0	0.31	0.31
	4	5	12.56	4.55	1	1.07	0.00	2	2.03	0.00	0	0.33	0.33
	6	0	9.25	9.25	1	1.61	0.23	1	2.90	1.25	1	0.30	1.62
	8	3	5.69	1.27	2	2.19	0.02	3	4.35	0.42	0	0.19	0.19
	10	0	2.62	2.62	0	1.18	1.18	1	2.21	0.66	0	0.13	0.13
	12	0	2.73	2.73	1	1.44	0.13	3	2.28	0.23	0	0.04	0.04
2	2	20	13.17	3.54	1	0.66	0.17	3	1.29	2.28	1	0.51	0.46
	4	34	15.73	21.20	4	1.57	3.77	2	2.33	0.05	0	0.45	0.45
	6	8	12.26	1.48	7	2.28	9.76	2	3.84	0.88	0	0.42	0.42
	8	6	6.75	0.08	1	3.12	1.44	9	5.97	1.54	0	0.24	0.24
	10	1	2.84	1.19	0	1.49	1.49	2	2.94	0.30	0	0.08	0.08
	12	1	3.53	1.82	1	1.78	0.34	13	3.01	33.21	2	0.12	30.47
4	2	2	15.41	11.67	0	0.89	0.89	1	1.36	0.10	0	0.49	0.49
	4	24	17.97	2.02	4	1.38	5.00	3	2.66	0.04	0	0.47	0.47
	6	14	13.70	0.01	5	2.53	2.42	6	4.19	0.78	0	0.52	0.52
	8	5	7.91	1.07	5	3.75	0.41	3	6.64	2.00	1	0.29	1.76
	10	0	4.09	4.09	1	1.87	0.40	1	3.07	1.39	0	0.12	0.12
	12	0	3.94	3.94	2	1.90	0.01	11	3.47	16.38	0	0.12	0.12
6	2	1	21.28	19.32	0	1.00	1.00	0	1.98	1.98	0	0.68	0.68
	4	5	25.27	16.26	1	2.26	0.70	0	4.12	4.12	0	0.82	0.82
	6	10	18.32	3.78	3	3.19	0.01	4	6.58	1.01	0	0.60	0.60
	8	5	11.10	3.36	1	5.13	3.33	1	8.72	6.83	0	0.36	0.36
	10	2	5.40	2.14	1	2.60	0.98	1	4.42	2.64	0	0.13	0.13
	12	1	5.55	3.73	1	2.67	1.05	5	4.69	0.02	0	0.16	0.16
8	2	1	24.65	22.69	0	1.38	1.38	0	2.12	2.12	0	0.73	0.73
	4	2	29.80	25.94	1	2.53	0.93	0	4.33	4.33	0	0.77	0.77
	6	0	22.53	22.53	3	3.94	0.22	0	7.57	7.57	0	0.69	0.69
	8	5	13.11	5.02	2	5.53	2.26	0	10.80	10.80	0	0.39	0.39
	10	3	5.89	1.42	0	2.73	2.73	1	5.42	3.60	0	0.21	0.21
	12	0	6.43	6.43	0	3.20	3.20	0	5.60	5.60	0	0.19	0.19

Table 5.13, continued

b. surface form = /hi/

		mod			wg			cr			gs		
init pitch	pitch span	obs. freq.	exp. freq.	χ^2	obs. freq.	exp. freq.	χ^2	obs. freq.	exp. freq.	χ^2	obs. freq.	exp. freq.	χ^2
-2	2	2	32.11	28.24	0	0.40	0.40	1	0.38	1.03	0	0.05	0.05
	4	8	26.40	12.82	0	0.64	0.64	0	0.88	0.88	0	0.01	0.01
	6	2	14.64	10.91	0	0.76	0.76	0	1.12	1.12	0	0.02	0.02
	8	1	4.91	3.12	0	0.75	0.75	0	1.01	1.01	0	0.00	0.00
	10	1	2.14	0.61	0	0.48	0.48	0	0.67	0.67	0	0.00	0.00
	12	2	1.87	0.01	0	0.39	0.39	0	0.60	0.60	0	0.00	0.00
0	2	30	60.20	15.15	0	0.72	0.72	0	0.97	0.97	0	0.10	0.10
	4	30	50.68	8.44	0	1.10	1.10	0	1.28	1.28	0	0.06	0.06
	6	5	27.90	18.80	0	1.70	1.70	0	2.16	2.16	0	0.01	0.01
	8	6	9.60	1.35	0	1.74	1.74	0	2.05	2.05	0	0.01	0.01
	10	0	3.95	3.95	0	0.82	0.82	0	1.10	1.10	0	0.01	0.01
	12	3	3.70	0.13	0	0.66	0.66	0	1.10	1.10	0	0.00	0.00
2	2	51	48.18	0.17	1	0.50	0.50	0	0.65	0.65	0	0.05	0.05
	4	56	40.70	5.75	0	0.85	0.85	0	1.13	1.13	0	0.02	0.02
	6	17	21.75	1.04	0	1.34	1.34	0	1.84	1.84	0	0.01	0.01
	8	6	7.23	0.21	0	1.23	1.23	0	1.61	1.61	0	0.01	0.01
	10	0	3.10	3.10	0	0.69	0.69	0	0.91	0.91	0	0.00	0.00
	12	3	3.06	0.00	1	0.70	0.13	1	0.95	0.00	0	0.00	0.00
4	2	13	43.55	21.43	0	0.45	0.45	0	0.69	0.69	0	0.06	0.06
	4	28	37.17	2.26	0	0.77	0.77	0	1.05	1.05	0	0.05	0.05
	6	33	19.41	9.52	0	1.05	1.05	0	1.52	1.52	0	0.00	0.00
	8	9	6.59	0.88	0	1.30	1.30	0	1.58	1.58	0	0.01	0.01
	10	1	2.80	1.16	1	0.60	0.27	0	0.73	0.73	0	0.00	0.00
	12	0	2.66	2.66	1	0.58	0.30	2	0.78	1.90	0	0.00	0.00
6	2	0	31.78	31.78	0	0.36	0.36	0	0.40	0.40	0	0.06	0.06
	4	6	26.26	15.63	0	0.59	0.59	0	0.87	0.87	0	0.02	0.02
	6	10	13.95	1.12	0	0.82	0.82	0	1.16	1.16	0	0.01	0.01
	8	6	4.84	0.28	0	0.79	0.79	0	1.20	1.20	0	0.01	0.01
	10	2	1.74	0.04	0	0.49	0.49	0	0.54	0.54	0	0.00	0.00
	12	0	2.00	2.00	0	0.46	0.46	1	0.55	0.37	0	0.00	0.00
8	2	0	26.08	26.08	0	0.23	0.23	0	0.49	0.49	0	0.03	0.03
	4	0	21.58	21.58	0	0.44	0.44	0	0.66	0.66	0	0.00	0.00
	6	3	12.62	7.34	0	0.62	0.62	0	0.94	0.94	0	0.02	0.02
	8	3	3.99	0.25	0	0.69	0.69	0	0.95	0.95	0	0.00	0.00
	10	0	1.75	1.75	0	0.31	0.31	0	0.47	0.47	0	0.01	0.01
	12	2	1.58	0.11	0	0.45	0.45	0	0.49	0.49	0	0.00	0.00

5.6 Yucatec Maya Grammar 2: Learning Perception and Production Simultaneously

In the last section, we saw that learning by perception tableaus alone leads to a production grammar with problematic areas. This suggests that the learner also has to test the relations among the constraint rankings for a given surface form (i.e. production learning). In this section we will add an additional learning step to the procedure in order to train the learner on the necessary data.

5.6.1 Modifying the Learning Strategy

In this new learning simulation, I start with the mean ranking values for the cue constraints that were obtained from the first learning simulation. This method allows for perception learning to happen first, as has been claimed in the literature. Thus the "initial" state of our current learning simulation does not have every constraint ranked at 100; the cue constraints start with the mean ranking values that were determined by the first learning step and the articulatory constraints start at 100 based on the assumption that their ranking values have not yet been learned. We know that we need to train the learner with production tableaus so that they can learn the relations among cue constraints in the /gl/ and /hi/ sets and so that they can learn the rankings of the articulatory constraints. It also seems likely that the learner would continue to learn through perception tableaus; we do not want a learning step that overwrites all perception learning in favor of production learning. For this reason, I include both perception and production tableaus in this learning step. At each point of evaluation, the learner checks to see if their current grammar predicts both the correct surface form to phonetic form mapping and the correct phonetic form to surface form mapping.

The perception tableaus (phonetic form to surface form mapping) are the same as those used in the first learning step. The only new tableaus in the model are the production tableaus. Recall from our discussion of lexicon-driven perception learning that different results were predicted to occur for a learner who used perception tableaus for all phonetic dimensions at once as compared to a learner who used perception tableaus for each phonetic dimension individually. The same is not true for production tableaus; the same ranking adjustments occur regardless of whether the output candidates are defined by all three phonetic dimensions or whether they are individually defined for each phonetic dimension. This fact is explained in more detail in §6.1.2. To simplify the files that are given to PRAAT in order to run the learning simulation, I defined outputs by all three phonetic dimensions at once.

This means that, in addition to training on the same perception tableaus, the learner at this stage is also trained on two different production tableaus: one with an input of a GLOTTALIZED vowel and one with an input of a HIGH TONE vowel. All possible combinations of glottalization type, initial pitch, and pitch span form the candidate set that is evaluated for each tableau. The distribution of the phonetic productions that occur with each of the surface forms is defined by the data from the production study.

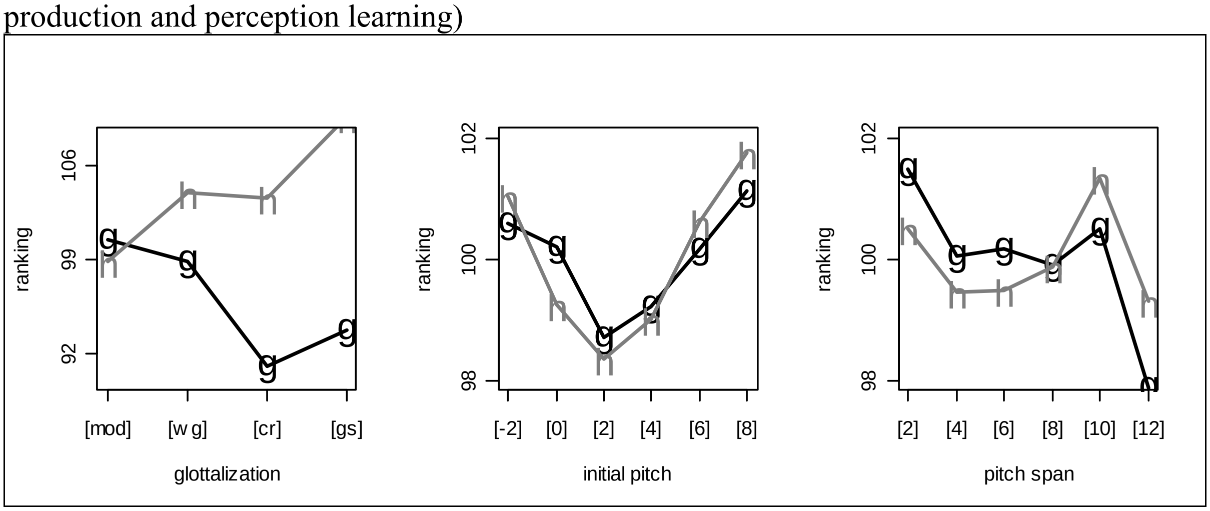
5.6.2 Results of the Learning Simulation

The grammar developed in this section started with cue constraints that have mean ranking values that were determined by perception learning. A new learning simulation is run (all PRAAT defaults) where the learner is simultaneously trained on both production and perception tableaus. The resulting mean ranking values are given in Table 5.14 and the rankings of the cue constraints are shown in Fig. 5.13.

Table 5.14: Mean ranking values of cue and articulatory constraints (developed through simultaneous production and perception learning)

	constraint	ranking value	constraint	ranking value
glottalization	*/gl/,[mod]	100.51	*/hi/,[mod]	98.78
	*/gl/,[wg]	99.00	*/hi/,[wg]	103.75
	*/gl/,[cr]	91.49	*/hi/,[cr]	103.20
	*/gl/,[gs]	93.91	*/hi/,[gs]	109.32
initial pitch	*/gl/,[in = -2]	100.44	*/hi/,[in = -2]	101.20
	*/gl/,[in = 0]	100.03	*/hi/,[in = 0]	99.38
	*/gl/,[in = 2]	98.59	*/hi/,[in = 2]	98.45
	*/gl/,[in = 4]	99.05	*/hi/,[in = 4]	99.13
	*/gl/,[in = 6]	100.01	*/hi/,[in = 6]	100.74
	*/gl/,[in = 8]	100.98	*/hi/,[in = 8]	101.93
pitch span	*/gl/,[sp = 2]	101.45	*/hi/,[sp = 2]	100.53
	*/gl/,[sp = 4]	99.99	*/hi/,[sp = 4]	99.56
	*/gl/,[sp = 6]	99.99	*/hi/,[sp = 6]	99.69
	*/gl/,[sp = 8]	99.66	*/hi/,[sp = 8]	100.14
	*/gl/,[sp = 10]	100.13	*/hi/,[sp = 10]	101.64
	*/gl/,[sp = 12]	97.27	*/hi/,[sp = 12]	99.87
articulatory	*[adducted vocal folds]		*[wg]	100.69
			*[cr]	97.93
			*[gs]	103.23
	*[creaky voice, small pitch span]		*[cr, sp<6]	99.22
			*[cr, sp<4]	100.33
			*[cr, sp<2]	98.96
	*[large pitch span]		*[sp>4]	98.00
			*[sp>6]	98.44
			*[sp>8]	98.74
			*[sp>10]	98.93
			*[sp>12]	97.15

Figure 5.13: Mean ranking values of cue constraints (developed through simultaneous production and perception learning)

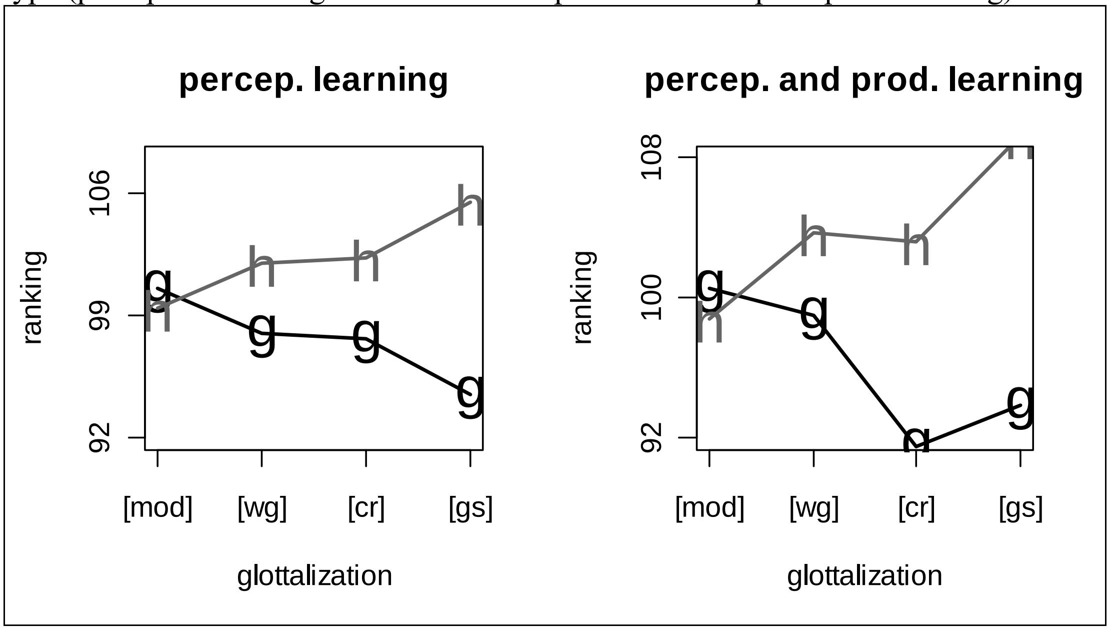


5.6.3 Assessment of the Grammar

In this section, I first discuss the generalizations that we can make from the figures, and then I present the predicted output distributions of both the production and perception grammars.

In Fig. 5.14 we see how the ranking values of the cue constraints that regulate glottalization type have changed when we add the second learning step. The constraints that penalize various forms of glottalization are now more lowly ranked for an input of a GLOTTALIZED vowel and more highly ranked for an input of HIGH TONE vowel. This is the result of having the articulatory constraints *[wg], *[cr], and *[gs] in the model. Recall that glottalization is a good cue to the GLOTTALIZED vowel in perception but that it is regularly not produced for the GLOTTALIZED vowel because it is effortful. With these articulatory constraints in the model, the cue constraints do not have to do the work of ensuring that glottalization is not always produced with the GLOTTALIZED vowel. Thus, */gl/,[gs] (for example) can have a low mean ranking value; this tells the perception grammar that [gs] is a good cue for perceiving a GLOTTALIZED vowel, and it tells the production grammar that /gl/ should be mapped onto [gs]. This mapping does not happen often in production, though, because *[gs] rules it out. This is essentially the 'prototype effect' (as discussed in Boersma 2006; see also Boersma and Hamann 2008); listeners prefer for GLOTTALIZED vowels to have glottal stops (as evidenced by the results of both perception tasks, Chapter 3) but this articulation is only marginally produced (as evidenced by the results of the production studies, Chapter 2).

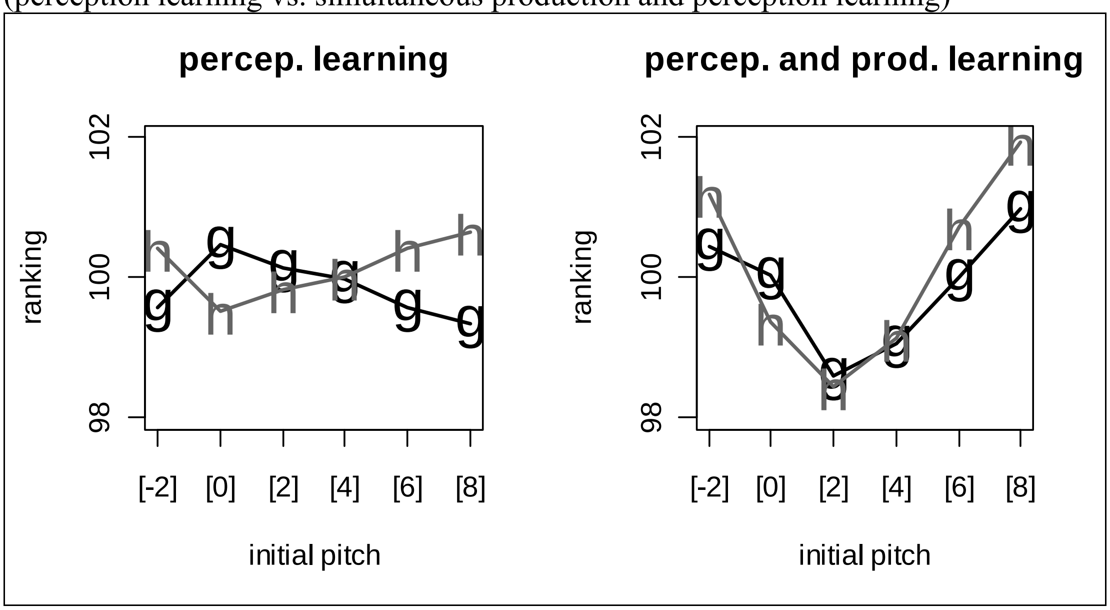
Figure 5.14: Comparison of the ranking values of cue constraints that regulate glottalization type (perception learning vs. simultaneous production and perception learning)



This constraint ranking predicts that a phonetic form with creaky voice or a glottal stop will almost certainly be mapped onto a GLOTTALIZED vowel. The predictions with regard to the production grammar are quite different with the new grammar because of the mean ranking values of the relevant articulatory constraints. For example, *[gs] has a mean ranking value of 103.23 (Table 5.14), and so this constraint is likely to assign the fatal violation mark to any candidates with a full glottal stop (except on the rare occasions where the evaluation ranking value of *[gs] is lower than the evaluation ranking values of */gl/,[mod]; */gl/,[cr]; or */gl/,[wg]). In general, the two different learning strategies do not result in drastically different ranking values for the cue constraints that regulate glottalization type, but different output distributions are predicted for the production grammar because of the inclusion of articulatory constraints.

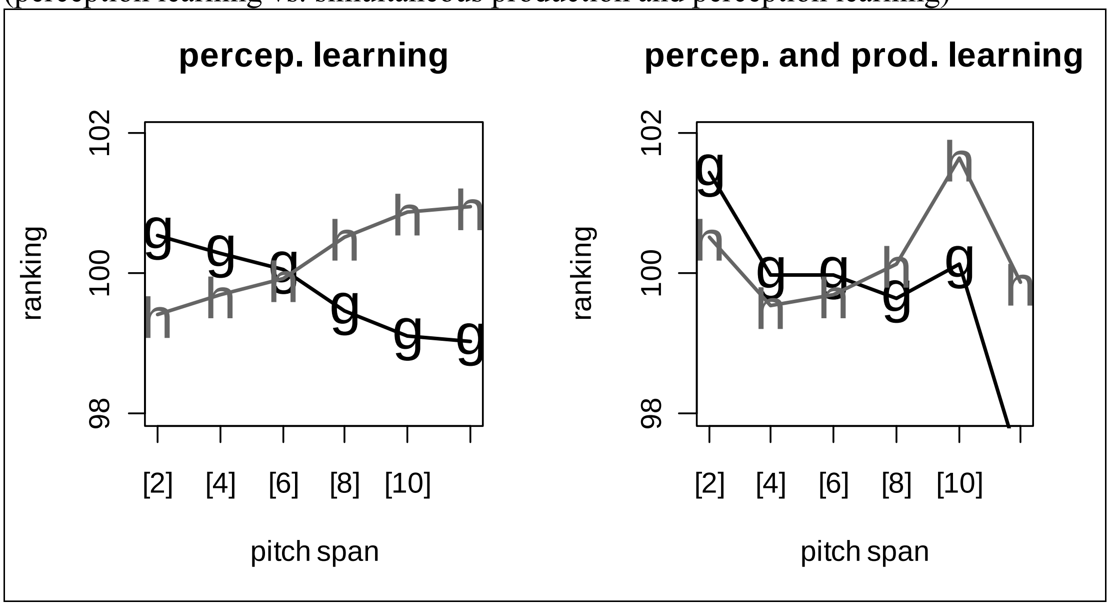
The two graphs in Fig. 5.15, which show the mean ranking values of the cue constraints that regulate initial pitch, are strikingly different. Recall from our discussion in the preceding section that the problem with perception learning was that the resulting grammar predicted too many productions of HIGH TONE vowels to have an initial pitch value of [0]. The new grammar now predicts for both GLOTTALIZED vowels and HIGH TONE vowels to be most often produced with an initial pitch of [2], which they are in the actual data. Thus, production learning has taught the learner how to rank the constraints in the /gl/ set and the /hi/ set with respect to each other. Furthermore, this new grammar still predicts for higher initial pitch values to be more likely to be mapped onto GLOTTALIZED vowels, which is again what we want. Because our grammar does not include any articulatory constraints that regulate initial pitch, the predicted output distributions (with regard to initial pitch) for both the production and perception data are determined solely from the ranking of the cue constraints.

Figure 5.15: Comparison of the ranking values of cue constraints that regulate initial pitch (perception learning vs. simultaneous production and perception learning)



In terms of pitch span, the two grammars are also noticeably different, as shown in Fig. 5.16. The mean ranking values of the cue constraints in the new grammar show that, in production, GLOTTALIZED vowels prefer to be mapped onto an extremely large pitch span. Of course we now have articulatory constraints in the model that can prevent this from happening too often, as a pitch span of [12] is not the most frequent pitch span in the data. This is similar to the discussion of glottalization above; large pitch spans are good cues to GLOTTALIZED vowels, but they are effortful and so articulatory constraints rule them out in production. With regard to HIGH TONE vowels, the production grammar now indicates that [2] and [4] should be the preferred values for pitch spans, and they are. The production grammar has thus been improved by learning through production tableaus.

Figure 5.16: Comparison of the ranking values of cue constraints that regulate pitch span (perception learning vs. simultaneous production and perception learning)



Even though the new grammar is clearly different from the first grammar, the perception grammars make about the same predictions. In both grammars, lower pitch spans are more likely to be mapped onto HIGH TONE vowels, and larger pitch spans are more likely to be mapped onto GLOTTALIZED vowels.

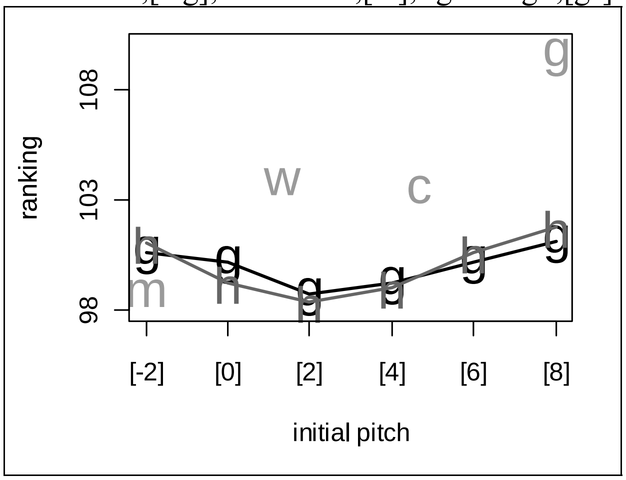
The output distributions that are predicted to occur with the new perception grammar are displayed in Table 5.15. If we first focus on the those tokens with modal voice, our perception grammar predicts roughly the same distributions as the first grammar: initial pitch values of [-2], [6], and [8] trigger more GLOTTALIZED vowel responses, and in general larger pitch spans trigger more GLOTTALIZED vowel responses. However, there is an anomaly in that a pitch span of [10] is associated with more GLOTTALIZED vowel responses than a pitch span of [12]. There are other anomalies in the weak glottalization and creaky voice categories. In these categories, initial pitch and pitch span do not have much of an effect on the participants' responses. There is a small increase in GLOTTALIZED vowel responses as the pitch span increases. Surprisingly, initial pitch categories of [2] and [4] trigger slightly more GLOTTALIZED vowel responses, even though these values should trigger more HIGH TONE vowel response than the peripheral initial pitch values.

Table 5.15: Percentage of times the perception grammar predicts a GLOTTALIZED vowel to be the optimal output for a given phonetic input (after simultaneous production and perception learning)

	pitch span	initial pitch -2	0	2	4	6	8
modal	2	43	31	32	33	40	47
	4	47	33	34	36	44	52
	6	47	34	35	37	45	52
	8	51	40	41	43	50	56
	10	60	53	56	57	60	62
	12	55	43	47	48	54	58
weak glottalization	2	79	78	80	80	79	79
	4	86	85	88	88	87	85
	6	86	85	88	88	87	85
	8	87	87	90	90	88	86
	10	88	88	91	90	89	87
	12	90	90	94	93	91	88
creaky	2	75	74	76	76	76	75
	4	84	83	87	86	85	82
	6	84	83	87	87	85	83
	8	86	85	90	89	87	84
	10	87	86	90	89	88	85
	12	89	90	96	94	91	87
glottal stop	2	100	100	100	100	100	100
	4	100	100	100	100	100	100
	6	100	100	100	100	100	100
	8	100	100	100	100	100	100
	10	100	100	100	100	100	100
	12	100	100	100	100	100	100

In order to determine why the initial pitch values of [2] and [4] are the most likely to be mapped onto GLOTTALIZED vowels with tokens that are produced with either weak glottalization or creaky voice, it will be helpful to reexamine the ranking values of the cue constraints that regulate initial pitch and glottalization. In Fig. 5.17 we see the mean ranking values of all the cue constraints that regulate initial pitch as well as the mean ranking values of the cue constraints that penalize the pairing of /hi/ with each type of glottalization (for such constraints the positioning of the marker on the x axis is irrelevant). What this graph shows is that, relative to the ranking values of all the constraints that regulate initial pitch, the ranking values of */hi/,[wg]; */hi/,[cr]; and */hi/,[gs] are quite high. This is of course desirable because this means that any form with glottalization is unlikely to be mapped onto a HIGH TONE vowel. However, the exact ranking values of these constraints cause an unexpected interaction between pitch and glottalization. When peripheral initial pitch values are produced (e.g. [in = -2]) in combination with glottalization, */hi/,[wg]; */hi/,[cr]; or */hi/,[gs] are likely to rule out the mapping of such a form onto a HIGH TONE vowel. However, because the constraint that penalizes the pairing [in = -2] with /gl/ also has a fairly high ranking value, there is a chance that this constraint will dominate the constraints that penalizes the pairing of /hi/ and glottalization (as well as */hi/,[in = -2]) and hence that this phonetic form will be mapped onto a HIGH TONE vowel.

Figure 5.17: Ranking values of select cue constraints
'g' = cue constraints that regulate the pairing of /gl/ with each initial pitch value; 'h' = cue constraints that regulate the pairing of /hi/ with each initial pitch value; 'm' = */hi/,[mod]; 'w' = */hi/,[wg]; 'c' = */hi/,[cr]; 'g' = */gl/,[gs]



The constraints that penalize the pairing of the central initial pitch values (e.g. [in = 2]) with either /gl/ or /hi/ have much lower mean ranking values. This means that these constraints are rarely in a position to determine the winning candidate. A phonetic form of [wg, in = 2] will not be mapped onto a HIGH TONE vowel because of */hi/,[wg]; the ranking of the constraints that regulate initial pitch will not even come into play. Similarly, the ranking value of */hi/,[gs] is so high that any form with [glottal stop] cannot be mapped onto a HIGH TONE vowel regardless of what other phonetic features occur with the glottal stop.

This phenomenon could be called a *ceiling effect*. When the pairing of one value of a phonetic dimension and one surface form is so disprefered, this pairing will be ruled out without any influence from the other phonetic dimensions. It seems that, in this case, the ceiling effect has led to undesirable results because real listeners are influenced by initial pitch even when it occurs with weak glottalization or creaky voice. It is thus the case that after adding production tableaus to the learning process, the perception grammar that was developed is not quite as good as the perception grammar that was developed from perception learning alone. However, the new perception grammar still predicts most of the same tendencies of real listeners. In order to fully assess this new grammar, we need to see how well the production grammar can generate output distributions that match those in the data set.

In Table 5.16 we see the information that is used to calculate the χ^2 (as was discussed in §5.3 before Table 5.11). With this new grammar, some outputs were predicted to occur 0 times (out of 100,000),[67] which is a problem for an equation that requires us to divide by the predicted frequency. In order to be able to perform the necessary calculations, each predicted frequency of 0 was changed to .0000001, meaning these values are predicted to occur less than 2^{-7}% of the time. Given stochastic evaluation of constraint rankings, this is a realistic expectation.

[67] The presence of articulatory constraints accounts for predicted frequency counts of 0. Because the glottal stop is so highly marked in this grammar, it is predicted to be rarely produced for GLOTTALIZED vowels and never produced for HIGH TONE vowels, which is an accurate description of real productions.

Table 5.16: Comparison of observed and expected frequencies of input and output pairings (simultaneous production and perception learning)
χ^2 = 60,000,542.46; adjusted χ^2 = 550.46

a. surface form = /gl/

		mod			wg			cr			gs		
init pitch	pitch span	obs. freq.	exp. freq.	χ^2	obs. freq.	exp. freq.	χ^2	obs. freq.	exp. freq.	χ^2	obs. freq.	exp. freq.	χ^2
-2	2	4	6.69	1.08	0	0.60	0.60	2	0.92	1.25	0	0.30	0.30
	4	4	12.28	5.58	0	1.19	1.19	2	1.97	0.00	0	0.62	0.62
	6	1	6.76	4.91	0	1.93	1.93	1	2.97	1.31	0	0.09	0.09
	8	2	3.77	0.83	1	2.84	1.19	2	4.16	1.12	1	0.00	9999998
	10	0	0.75	0.75	1	0.63	0.22	0	1.08	1.08	0	0.00	0.00
	12	5	3.08	1.20	0	1.98	1.98	3	3.57	0.09	0	0.00	0.00
0	2	16	9.99	3.62	1	0.90	0.01	1	1.46	0.14	0	0.64	0.64
	4	5	18.21	9.59	1	1.58	0.21	2	2.84	0.25	0	0.79	0.79
	6	0	9.60	9.60	1	2.86	1.21	1	4.57	2.79	1	0.21	2.92
	8	3	5.37	1.05	2	3.77	0.83	3	6.42	1.82	0	0.00	0.00
	10	0	1.27	1.27	0	0.87	0.87	1	1.62	0.24	0	0.00	0.00
	12	0	4.67	4.67	1	3.27	1.57	3	5.30	1.00	0	0.00	0.00
2	2	20	29.14	2.87	1	2.38	0.80	3	4.38	0.44	1	1.76	0.33
	4	34	51.04	5.69	4	5.00	0.20	2	7.82	4.33	0	2.24	2.24
	6	8	28.25	14.51	7	7.98	0.12	2	13.52	9.81	0	0.54	0.54
	8	6	15.79	6.07	1	10.49	8.58	9	17.49	4.12	0	0.00	0.00
	10	1	3.98	2.23	0	2.60	2.60	2	4.53	1.42	0	0.00	0.00
	12	1	12.91	10.99	1	8.27	6.39	13	14.14	0.09	2	0.00	39999996
4	2	2	21.24	17.43	0	1.71	1.71	1	2.95	1.29	0	1.10	1.10
	4	24	36.79	4.45	4	3.58	0.05	3	5.80	1.35	0	1.51	1.51
	6	14	19.61	1.61	5	6.10	0.20	6	9.56	1.33	0	0.41	0.41
	8	5	11.02	3.29	5	7.27	0.71	3	12.65	7.36	1	0.00	9999998
	10	0	2.59	2.59	1	1.86	0.40	1	3.17	1.49	0	0.00	0.00
	12	0	8.75	8.75	2	6.49	3.11	11	10.49	0.02	0	0.00	0.00
6	2	1	10.19	8.29	0	0.92	0.92	0	1.38	1.38	0	0.55	0.55
	4	5	17.78	9.18	1	1.63	0.24	0	3.08	3.08	0	0.74	0.74
	6	10	10.17	0.00	3	3.06	0.00	4	4.48	0.05	0	0.21	0.21
	8	5	5.43	0.03	1	3.43	1.72	1	5.74	3.91	0	0.00	0.00
	10	2	1.33	0.34	1	0.90	0.01	1	1.59	0.22	0	0.00	0.00
	12	1	4.72	2.93	1	3.16	1.47	5	4.94	0.00	0	0.00	0.00
8	2	1	4.18	2.42	0	0.36	0.36	0	0.68	0.68	0	0.24	0.24
	4	2	7.62	4.15	1	0.79	0.05	0	1.17	1.17	0	0.30	0.30
	6	0	4.28	4.28	3	1.28	2.31	0	2.05	2.05	0	0.08	0.08
	8	5	2.13	3.87	2	1.45	0.21	0	2.75	2.75	0	0.00	0.00
	10	3	0.47	13.51	0	0.33	0.33	1	0.74	0.09	0	0.00	0.00
	12	0	1.97	1.97	0	1.36	1.36	0	2.06	2.06	0	0.00	0.00

Table 5.16, continued

b. surface form = /hi/

		mod			wg			cr gs					
init pitch	pitch span	obs. freq.	exp. freq.	χ^2	obs. freq.	exp. freq.	χ^2	obs. freq.	exp. freq.	χ^2	obs. freq.	exp. freq.	χ^2
-2	2	2	9.41	5.83	0	0.03	0.03	1	0.11	7.23	0	0.00	0.00
	4	8	12.02	1.35	0	0.09	0.09	0	0.20	0.20	0	0.00	0.00
	6	2	6.27	2.91	0	0.14	0.14	0	0.29	0.29	0	0.00	0.00
	8	1	2.73	1.10	0	0.12	0.12	0	0.34	0.34	0	0.00	0.00
	10	1	0.36	1.12	0	0.04	0.04	0	0.05	0.05	0	0.00	0.00
	12	2	1.27	0.41	0	0.05	0.05	0	0.21	0.21	0	0.00	0.00
0	2	30	41.29	3.09	0	0.29	0.29	0	0.45	0.45	0	0.00	0.00
	4	30	54.34	10.90	0	0.47	0.47	0	0.66	0.66	0	0.00	0.00
	6	5	28.10	18.99	0	0.95	0.95	0	1.42	1.42	0	0.00	0.00
	8	6	11.07	2.32	0	0.76	0.76	0	1.21	1.21	0	0.00	0.00
	10	0	2.06	2.06	0	0.18	0.18	0	0.24	0.24	0	0.00	0.00
	12	3	5.80	1.35	0	0.49	0.49	0	0.86	0.86	0	0.00	0.00
2	2	51	74.96	7.66	1	0.49	0.52	0	0.75	0.75	0	0.00	0.00
	4	56	96.22	16.81	0	0.83	0.83	0	1.20	1.20	0	0.00	0.00
	6	17	50.15	21.91	0	1.37	1.37	0	2.37	2.37	0	0.00	0.00
	8	6	19.29	9.16	0	1.38	1.38	0	2.20	2.20	0	0.00	0.00
	10	0	3.05	3.05	0	0.21	0.21	0	0.48	0.48	0	0.00	0.00
	12	3	10.38	5.25	1	0.87	0.02	1	1.29	0.06	0	0.00	0.00
4	2	13	46.93	24.53	0	0.23	0.23	0	0.51	0.51	0	0.00	0.00
	4	28	61.29	18.08	0	0.51	0.51	0	1.04	1.04	0	0.00	0.00
	6	33	31.26	0.10	0	1.08	1.08	0	1.32	1.32	0	0.00	0.00
	8	9	12.65	1.05	0	0.79	0.79	0	1.39	1.39	0	0.00	0.00
	10	1	2.01	0.51	1	0.16	4.50	0	0.30	0.30	0	0.00	0.00
	12	0	6.67	6.67	1	0.55	0.36	2	0.91	1.30	0	0.00	0.00
6	2	0	12.95	12.95	0	0.05	0.05	0	0.16	0.16	0	0.00	0.00
	4	6	17.84	7.86	0	0.14	0.14	0	0.20	0.20	0	0.00	0.00
	6	10	9.34	0.05	0	0.20	0.20	0	0.47	0.47	0	0.00	0.00
	8	6	3.82	1.25	0	0.22	0.22	0	0.31	0.31	0	0.00	0.00
	10	2	0.67	2.63	0	0.05	0.05	0	0.11	0.11	0	0.00	0.00
	12	0	1.82	1.82	0	0.12	0.12	1	0.34	1.31	0	0.00	0.00
8	2	0	3.88	3.88	0	0.04	0.04	0	0.05	0.05	0	0.00	0.00
	4	0	5.32	5.32	0	0.03	0.03	0	0.08	0.08	0	0.00	0.00
	6	3	2.82	0.01	0	0.10	0.10	0	0.15	0.15	0	0.00	0.00
	8	3	1.17	2.85	0	0.06	0.06	0	0.14	0.14	0	0.00	0.00
	10	0	0.16	0.16	0	0.03	0.03	0	0.03	0.03	0	0.00	0.00
	12	2	0.59	3.38	0	0.05	0.05	0	0.02	0.02	0	0.00	0.00

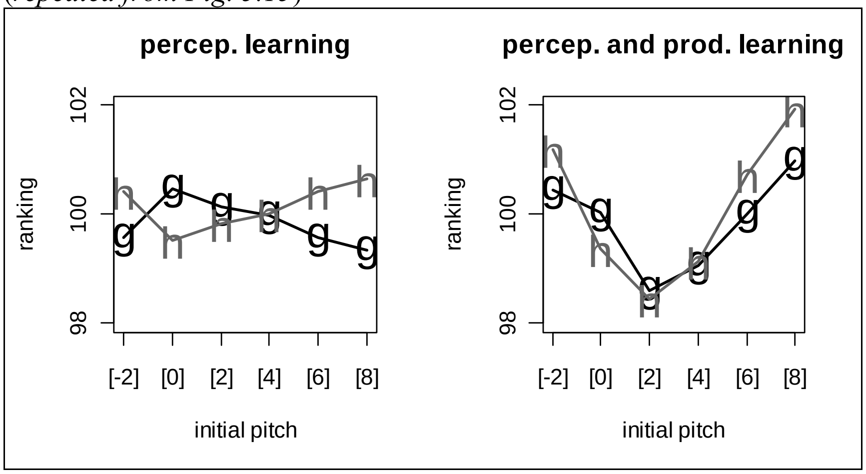
The χ^2 statistic for this production grammar is 60,000,542, while the χ^2 statistic for the first production grammar (after learning the rankings of articulatory constraints) is 835.74. As larger numbers imply a great degree of divergence, it seems that our new production grammar does even worse at predicting the output distributions of Yucatec Maya speech. However, if we take a closer look at Table 5.14, we see that there are 3 cells where our grammar predicted frequency counts of 0 (which were changed to 0.0000001 in order to calculate the χ^2 statistic) but where either one or two tokens were actually produced. The tokens [gs, in = -2, sp = 8], [gs, in = 2, sp = 12], and [gs, in = 4, sp = 8] are expected to not be observed, but they were observed (once, twice, and once respectively). When we consider how the terms of the χ^2 statistic are calculated – (obs. freq. – exp. freq.)^2/exp. freq. – we see that this situation causes division by a miniscule number (.0000001), resulting in a very large number. The total for these three cells is 59,999,992, and so if we remove them from our χ^2 calculation, we get a value of 550.46, which represents a lower degree of divergence than grammar 1.

The question is then whether this new grammar is better (as suggested by the adjusted χ^2 of 550) or worse (as suggested by the χ^2 of 60 million). One way to make the comparison between the χ^2 statistics of each grammar more appropriate is to remove the same cells from the first calculation, i.e. to subtract the χ^2 values associated with the productions [gs, in = -2, sp = 8], [gs, in = 2, sp = 12], and [gs, in = 4, sp = 8]. If we do this, the adjusted χ^2 statistic for grammar 1 would only drop by 37 – to 791. We could instead remove the three highest values, which would result in χ^2 = 730, which is still much smaller than 60 million, but larger than 550. It is clear that the new grammar (developed through simultaneous production and perception learning) is more problematic for three and only three phonetic outputs, whereas the first grammar (developed through perception learning for cue constraints and production learning for articulatory constraints) has more trouble with many different outputs. Another way to look at the data is to count the number of cells in which one grammar is more accurate than that other grammar: grammar 1 is more accurate than grammar 2 in 85 cells, while grammar 2 is more accurate than grammar 1 in 189 cells. Again, the grammar developed through perception learning of cue constraints and production learning of articulatory constraints (grammar 1) has multiple systematic errors, while the grammar developed through simultaneous production and perception learning has a few (extremely large) localized errors.

It is my conclusion that, in comparing the accuracy of these two production grammars, the false prediction of a frequency count of 0 in three cells is not as important as the fact that the overall ranking of the cue constraints as developed through simultaneous production and perception learning is able to capture more of the subtle patterns in the production data than the grammar that was developed through production learning alone. This is best illustrated by the ranking of the constraints that regulate initial pitch, as shown in Fig. 5.18 below, repeated from Fig. 5.15). We know that the initial pitch values produced for HIGH TONE and GLOTTALIZED vowels have a large degree of overlap: most initial pitch values that are legitimate for HIGH TONE vowels are also legitimate for GLOTTALIZED vowels, even though the average initial pitch values of the two vowel shapes are significantly different. This fact is clearly represented by the right-hand graph in Fig. 5.18, where we see that [2] is good value for both vowel shapes, and that higher initial pitch values are (slightly) preferably paired with GLOTTALIZED vowels and that lower initial pitch values are (slightly) preferably paired with HIGH TONE vowels. The left-hand graph on the other hand shows that

GLOTTALIZED vowels should be most often produced with peripheral pitch values, which is not an accurate reflection of the facts. Production learning was necessary to create the asymmetrical rankings in the right-hand graph.[68]

Figure 5.18: Comparison of the ranking values of cue constraints that regulate initial pitch (perception learning vs. simultaneous production and perception learning) (*repeated from Fig. 5.15*)



When testing how the output distributions predicted by the grammar compare to those demonstrated by real speakers and listeners, it is clear that the simultaneous production and perception learning has led to the development of a slightly worse perception grammar and a much better production grammar. The anomalies in the perception grammar are that a pitch span of [10] is predicted to trigger more GLOTTALIZED vowel responses than a pitch span of [12], and that, when glottalization is produced, central initial pitch values are predicted to trigger more GLOTTALIZED vowel responses. The production grammar, on the other hand, is able to match the actual output distributions from the data set with a high degree of accuracy (if we are willing to ignore three problematic outputs).

The results of this learning simulation show that production learning is successful in

[68] Note that the ranking of the constraints */hi/,[in=6]; */hi/,[in=8]; */gl/,[in=6]; and */gl/,[in=8] in the right-hand graph in Fig. 5.18 look similar to the rankings of the constraints in our hypothetical grammar in Fig. 5.9. Specifically, we see the pattern that [in = 6] is more preferably produced for either vowel shape and that either initial pitch value is more likely mapped onto /gl/. This is exactly the kind of asymmetrical pattern that was introduced in the hypothetical language data in (5.5).

that it allows the learner to learn the relations of the constraints in the /gl/ set and the /hi/ set – something that cannot be done through perception learning alone. By learning these relations, the new production grammar is able to account for phenomena that the first production grammar could not. For example, the fact that [2] is the preferred initial pitch value for both HIGH TONE and GLOTTALIZED vowels could not be learned through perception learning but was learned when production tableau were added to the learning process.

Though we have lost a bit of perception accuracy, the inaccuracies of the new perception grammar are sporadic and do not seem to show the same type of systematic errors that the first production grammar did. It is thus a possibility that variations in the constraint set could fix these sporadic errors.

In conclusion, the first production grammar (learned through perception learning) was flawed in that it could not rule out particular unmarked values for initial pitch and pitch spans because this information could not be learned from perception tableaus alone. By adding production learning, this information can be and is learned. The new learning step has resulted in complications for the perception grammar, but these complications do not denote the same type of systematic problems that we saw in the first production grammar.

5.7 Mean Ranking Values of Articulatory Constraints

Tables 5.12 and 5.14 provided the mean ranking values of articulatory constraints as determined by the GLA (using two different learning strategies). The ranking values in Table 5.14 (where a learner was trained on both production and perception tableaus and where both cue and articulatory constraint values were adjusted in accord with the GLA) are repeated below in Table 5.17. The constraints in each group are ordered in terms of a stringency scale where the topmost constraint is the most general and the bottommost constraint is the least general. For example, *[wg] penalizes any form of glottalization (weak glottalization, creaky voice, and glottal stop) and hence assigns a violation mark to any candidate that incurs a violation of *[cr] and/or *[gs]. We see in Table 5.17 that the mean ranking values of the articulatory constraints do not follow the stringency ordering of the constraints (as was also the case with the articulatory constraints in Table 5.12). In fact, within each set there are anti-Paninian rankings, such that a more general constraint dominates a more specific constraint within a given stringency system (see Prince 1997a-b, de Lacy 2002); one such example is the fact the mean ranking value of *[wg] (the most general constraint in the *[adducted vocal folds] category) is higher than the mean ranking value of *[cr] (which only assigns violations marks to [creaky voice] and [glottal stop] and not to [weak glottalization]). To be clear, if there were no anti-Paninian rankings in the table, each constraint would have a higher mean ranking value than the one above it (within the same group).

Table 5.17: Mean ranking values of articulatory constraints (simultaneous production and perception learning)
(partially repeated from Table 5.14)

*[adducted vocal folds]	*[wg]	100.69
	*[cr]	97.93
	*[gs]	103.23
*[creaky voice, small pitch span]	*[cr, sp<6]	99.22
	*[cr, sp<4]	100.33
	*[cr, sp<2]	98.96
*[large pitch span]	*[sp>4]	98.00
	*[sp>6]	98.44
	*[sp>8]	98.74
	*[sp>10]	98.93
	*[sp>12]	97.15

It is thus the case that the mean ranking values for articulatory constraints that the GLA has determined to be optimal in accounting for the Yucatec Maya data do not follow a hierarchical ranking, such as *[gs] » *[cr] » *[wg], where constraints are ordered from least to most general. As discussed in de Lacy (2002), anti-Paninian rankings can account for phenomena that fixed hierarchical rankings cannot. Specifically, anti-Paninian rankings can lead to category conflation. For example, the constraint *[wg] conflates all categories of glottalization; [wg], [cr], and [gs] each receive one violation mark from this constraint. Hence, this constraint cannot decide among these candidates, and some lower ranking constraint(s) must determine the winning candidate. In this section I show why such category conflation is ideal for certain parts of the Yucatec Maya data.

The tableau in (5.8) shows how the production grammar, where each constraint is ranked according to its mean ranking value (see Table 5.14), maps an input of /gl/ onto [modal voice]. We know that about half of GLOTTALIZED vowels are produced with modal voice, so this is a desirable result.

(5.8) production tableau where each constraint has an evaluation ranking value equal to its mean ranking value

/gl/	*[gs]	*[wg]	*/gl/,[mod]	*/gl/,[wg]	*[cr]	*/gl/,[gs]	*/gl/,[cr]
☞ [mod]			*				
[wg]		*!		*			
[cr]		*!			*		*
[gs]	*!	*			*	*	

Because of stochastic evaluation, we know that the ranking shown in (5.8) is not the ranking that will be used at every point of evaluation. If the evaluation ranking values of *[wg] and */gl/,[mod] are such that */gl/,[mod] » *[wg], then [mod] will no longer be the winning candidate, as shown in (5.9). Because *[wg] conflates all of the glottalization categories, the winning candidate is determined by the ranking of */gl/,[wg] and *[cr], as indicated by the dashed line in (5.9). In this way, a high-ranking *[wg] will rule out any form of glottalization as long as it dominates */gl/,[mod] at that point of evaluation. When the reverse ranking occurs, *[wg] cannot determine the winner, and hence other lower-ranking constraints become active and determine which category of glottalization is optimal.

(5.9) production tableau where */gl/,[mod] » *[wg] and the winning candidate is determined by the ranking of */gl/,[wg] and *[cr]

/gl/	*[gs]	*/gl/,[mod]	*[wg]	*/gl/,[wg]	*[cr]	*/gl/,[gs]	*/gl/,[cr]
[mod]		*!					
☞ [wg]			*	*(!)			
☞ [cr]			*		*(!)		*
[gs]	*!		*		*	*	

If we remind ourselves of the distribution of glottalization types in the Yucatec Maya data, it is easy to see why the ranking of *[wg] is so crucial. In the data that the simulated learner was trained on, [mod] occurs 56% of the time, [wg] occurs 15% of the time, [cr] occurs 26% of the time, and [gs] occurs 2% of the time (for GLOTTALIZED vowels). As stated above, the general *[wg] is able rule out any form of glottalization as long as it dominates */gl/,[mod], which it will more than half the time according to the mean ranking values of these constraints. When the ranking */gl/,[mod] » *[wg] occurs, the decision between [wg] and [cr] is left to */gl/,[wg] and *[cr]. Stochastic evaluation predicts for the ranking */gl/,[wg] » *[cr] which would lead to the selection of [cr] as the winning candidate) to occur more often than the ranking *[cr] » */gl/,[wg] (which would lead to the selection of [wg] as the winning candidate).

If we use a system with fixed hierarchical rankings such that, when ordered by mean ranking values, *[gs] » *[cr] » *[wg], the grammar that can account for the same patterns in production has some undesirable consequences for perception. The tableau in (5.10) uses the same constraint ranking as in (5.8) except that *[wg] is demoted below *[cr] in order to maintain the fixed ranking. In this tableau it is the ranking of */gl/,[mod], */gl/,[wg], and *[cr] that determines the winning candidate. We see that, in order for [mod] to win, */gl/,[mod] must be dominated by */gl/,[wg] and *[cr]. Recall that we want [mod] to win

56% of the time. This means that, in a system where *[cr] » *[wg], the mean ranking value of */gl/,[mod] must be lower and/or the mean ranking values of */gl/,[wg] and *[cr] must be higher. Such ranking values have undesirable consequences for the perception grammar because, in general, [modal voice] is not a good cue to GLOTTALIZED vowels (meaning we want */gl/,[mod] to be ranked fairly high) and [weak glottalization] is a good cue to GLOTTALIZED vowels (meaning we want */gl/,[wg] to be ranked fairly low). We can thus see how the anti-Paninian ranking of *[wg] » *[cr] is crucial in accounting for both the production and perception facts about Yucatec Maya.

(5.10) production tableau where anti-Paninian rankings are not allowed (*[gs] » *[cr] » *[wg])

/gl/	*[gs]	*/gl/, [mod]	*/gl/, [wg]	*[cr]	*[wg]	*/gl/, [gs]	*/gl/, [cr]
☞ [mod]		*(!)					
☞ [wg]			*(!)		*		
☞ [cr]				*(!)	*		*
[gs]	*!			*	*	*	

In order to test the differences between a grammar with freely rankable scalar constraints and one with fixed rankings, a learning simulation was run where the markedness constraints were set in a fixed hierarchical ranking. This learning simulation used the same format as that used to develop grammar 2 (see §5.6), where cue constraints start with mean ranking values as determined by perception learning, articulatory constraints start at 100, and the learner is trained on both production and perception tableaus. The resulting mean ranking values of the articulatory constraints are shown in Table 5.18 (ranking values of the cue constraints that result from this learning simulation are in Appendix B). This grammar was used to predict output distributions, and these output distributions were used to calculate the χ^2 statistic. The full table is in Appendix B; the χ^2 statistic 612.04 (adjusted χ^2 = 587.29), which is slightly higher than the adjusted χ^2 of 550.46 that we obtained from grammar 2. It is thus the case that fixed hierarchical rankings of articulatory constraints can lead to a grammar that is almost as accurate as one where articulatory constraints are freely rankable. However such a grammar is only *almost* as accurate and was not the grammar that was reached by a GLA learner who was not limited as to when to adjust the ranking values of articulatory constraints. Considering that the grammar that retains fixed hierarchical ranking values of articulatory constraints is not better than one with freely rankable articulatory constraints, I conclude that a GLA learner must be free to develop anti-Paninian rankings as these are sometimes necessary in accounting for language data.

Table 5.18: Mean ranking values of articulatory constraints

		no fixed rankings	fixed rankings
*[adducted vocal folds]	*[wg]	100.69	90.44
	*[cr]	97.93	95.82
	*[gs]	103.23	102.20
*[creaky voice, small pitch span]	*[cr, sp<6]	99.22	97.60
	*[cr, sp<4]	100.33	97.61
	*[cr, sp<2]	98.96	97.62
*[large pitch span]	*[sp>4]	98.00	77.06
	*[sp>6]	98.44	83.08
	*[sp>8]	98.74	87.64
	*[sp>10]	98.93	93.11
	*[sp>12]	97.15	99.06

The fact that anti-Paninian rankings of articulatory constraints are necessary in accounting for production in Yucatec Maya can be used to argue against the idea that articulatory constraints have fixed universal ranking or the idea that the only language-specific adjustments made to the rankings of articulatory constraints are those that lower the ranking value of constraints that penalized practiced articulations. If these constraints have fixed universal values, we would expect these values to follow the hierarchy of the stringency scale. There is certainly no reason to propose that, in all languages, *[wg] » *[cr]. It would also be unreasonable to suggest that the universal ranking is *[cr] » *[wg] and that, due to practice alone, the Yucatec Maya speaker ends up with *[wg] » *[cr]. Both [wg] and [cr] are almost equally common with Yucatec Maya GLOTTALIZED vowels, and so if practice reduces the ranking value of *[wg], it should also reduce the ranking value of *[cr]. Further theoretical consequences of these anti-Paninian rankings are discussed in §6.1.4.

It is thus clear that some type of production learning must be used in addition to perception learning in order to reach an accurate adult grammar of the phonetics-phonology interface. I propose that production learning is necessary for both articulatory and cue constraints. The fact that a GLA learner develops anti-Paninian rankings of articulatory constraints is inconsistent with the hypothesis that articulatory constraints have fixed cross-linguistic rankings. Furthermore, the use of production learning with cue constraints led to a more accurate production grammar, as discussed in §5.6.

5.8 Chapter Summary

The main conclusion of this chapter is that there are aspects of the production grammar that cannot be learned through perception learning only. The language learner must test their interim grammar with both production and perception tableaus in order to learn the relations among the mean ranking values of all of the cue constraints. Furthermore, production learning is also necessary for ranking the articulatory constraints, as the anti-Paninian rankings that best account for the Yucatec Maya data are clearly language-specific rankings.

In this chapter we have tested multiple learning strategies in the development of a grammar of Yucatec Maya that accounts for the production and perception of pitch and glottalization. The first learning strategy was the one used by Boersma (2006) and Boersma and Hamann (2008) – lexicon-driven perceptual learning (called simply perception learning here), which is used to learn the rankings of cue constraints. This strategy led to a highly accurate perception grammar. In order to fully evaluate the production grammar, articulatory constraints were added to the model. Though Boersma and Hamann (2008) propose that articulatory constraints have fixed universal rankings (where the only language specific adjustments to ranking values are that practiced articulations are deemed less effortful), the fact that there is no literature on what these ranking might be made it best to let the learner determine the rankings of articulatory constraints. For this reason, a learning simulation was run where the cue constraints maintained the ranking values that were determined through perceptual learning and production learning was used to rank the articulatory constraints. This resulted in a fairly accurate production grammar, but one that had systematic problems in that it was not able to rule out certain unmarked values that are not the preferred values for initial pitch and pitch span.

A new learning simulation was run in which production and perception learning are done simultaneously (and in which both production and perception tableaus affect the rankings of cue constraints). This simulation led to a highly accurate production grammar. Production learning led to asymmetrical rankings of cue constraints (see discussion in §5.3), which are especially necessary for the cue constraints that regulate initial pitch values. Because the GLOTTALIZED and HIGH TONE vowels are produced with similar distributions of initial pitch values, the appropriate adult grammar is one in which, for example, the constraints */hi/,[in = 2] and */gl/,[in=2] both have low mean ranking values. Perception learning cannot lead to this grammar because, which such a learning strategy, every time that */hi/,[in=2] is demoted, */gl/,[in=2] is promoted (and vice versa). In this way, production learning is necessary to demote both of these constraints.

The grammar that resulted from simultaneous production and perception learning (§5.6) captures most of the generalizations that we made the data from actual speakers and listeners of Yucatec Maya (§5.1). In production, GLOTTALIZED vowels are produced with all the types of glottalization and tend to be produced with higher initial pitch values and larger pitch spans, while HIGH TONE vowels are produced with modal voice and lower initial pitch values and smaller pitch spans. In perception, glottalization of any form is preferably mapped onto GLOTTALIZED vowels, while modal voice is preferably mapped onto HIGH TONE vowels. The phonetic dimensions of initial pitch and pitch span are also predicted to be cues to distinguishing this contrast, but they are not as robust as glottalization. However, when vowels are produced with modal voice, which could signal a production of either a GLOTTALIZED or a HIGH TONE vowel, initial pitch and pitch span are predicted to be used as cues by the listener in determining whether a GLOTTALIZED or HIGH TONE vowel was said.

6

Implications for Future Work with Bidirectional Stochastic OT and the GLA

In this chapter I go into further details about some of the subtleties of working with Bidirectional OT and the GLA. In §6.1 I look at the effects of applying different approaches to the model. These include the use of finely grained cue constraints, the differences between the evaluation of multiple phonetic dimensions at once versus individually in both production and perception, and the use of weighted constraints. In §6.2 I discuss both the synchronic and diachronic predictions made by using a Stochastic OT model and a learner that simultaneously evaluates both production and perception tableaus.

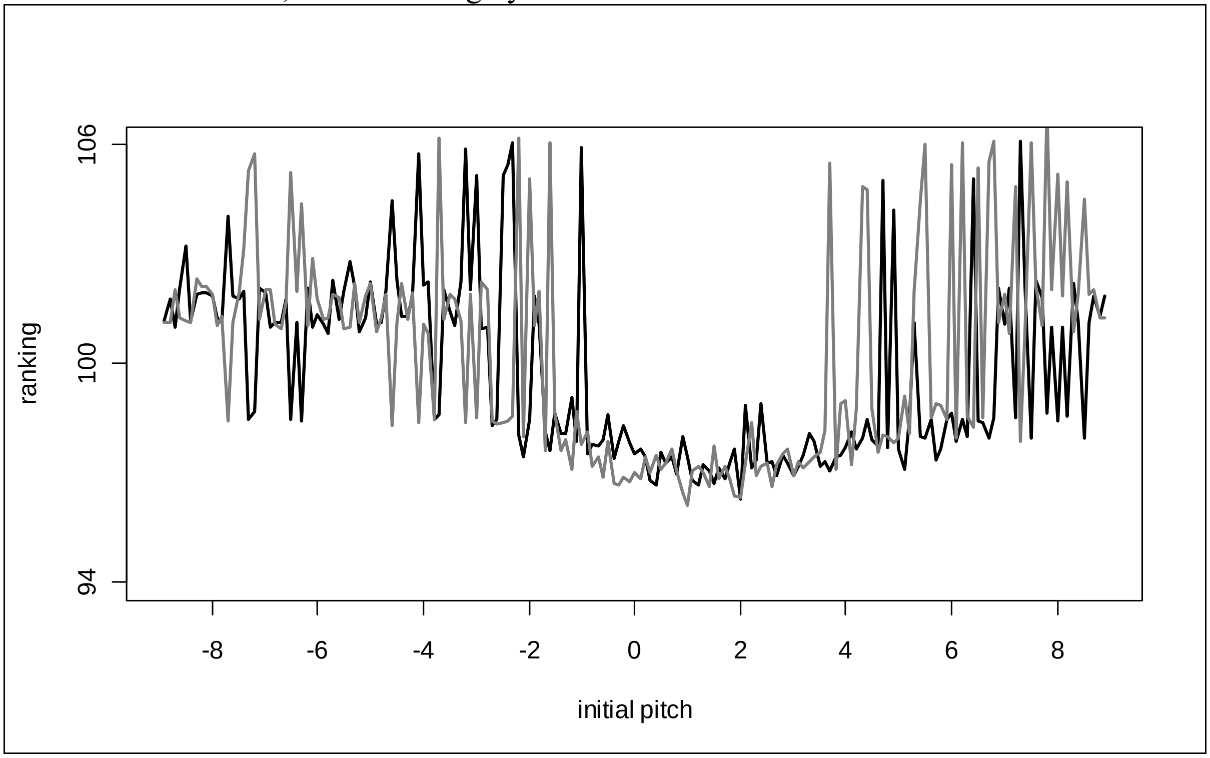
6.1 Different Approaches to the Model

6.1.1 Finely Grained Cue Constraints

In Chapter 5 we used cue and articulatory constraints that reference wider categories of phonetic dimensions than are desirable for the phonetics-phonology interface. The goal of this interface is to account for the relation between a discrete phonological form and a continuous form. That continuous form has to be categorized in some way in order to define a grammar with a finite set of constraints, but it is more appropriate to use very finely grained categories than to use the larger categories that I used in that chapter. As was explained at that time, a data set of 685 tokens is not enough information for the development of a meaningful grammar with smaller phonetic categories. In this section, we will explore the limitations of this data set and see how analysis with finely grained constraints shows promise for use with larger data sets. I hope for this discussion to provide the motivation for future work in collecting and analyzing larger data sets.

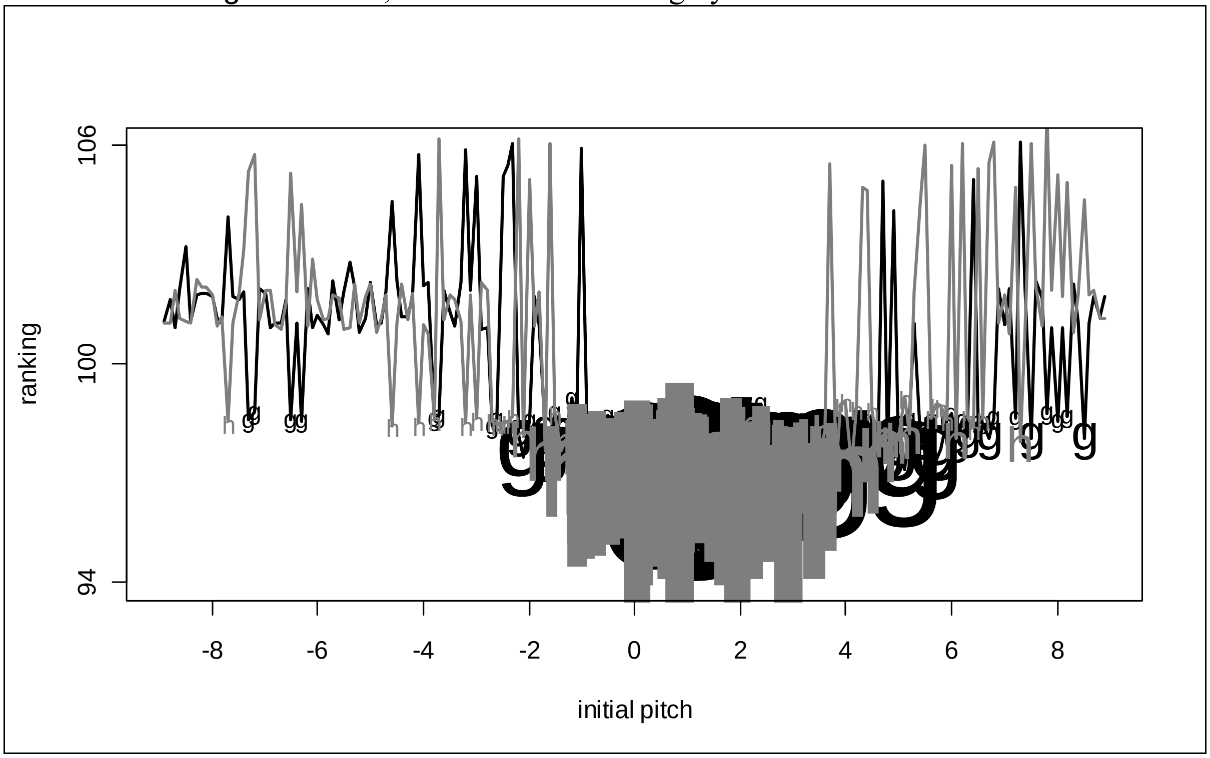
In this section we look at only the phonetic dimension of initial pitch, and we will categorize this dimension by 1/10 of a semitone. In Fig. 5.4 we saw a histogram of initial pitch values for the glottalized vowel, and this histogram showed that the data was skewed such that there are a few tokens with extremely low initial pitch values. I will exclude some of these low values and look only at initial pitch values between -8.9 s/b and 8.9 s/b. The grammar developed in this section is defined by cue constraints that penalize the pairing of each possible value for initial pitch (from -8.9 s/b to 8.9 s/b by increments of 0.1) with each of the two possible surface forms (/gl/ or /hi/). The distribution of the pairings of each initial pitch value with each surface form that the learner is trained on comes from the actual pairings that occur in the data set of 685 tokens, as before. A learning simulation is run with this information, using all PRAAT defaults. This simulation follows the strategy of simultaneous learning from production and perception tableaus, as was developed in §5.6. The resulting mean ranking values of the cue constraints are shown in Fig. 6.1.

Figure 6.1: Mean ranking values of cue constraints that regulate initial pitch
GLOTTALIZED = black; HIGH TONE = gray



The above figure does not allow us to make the same neat generalizations that we were able to make from previous graphs of the ranking values of cue constraints. If we focus on the constraints with lower mean ranking values (which denote preferred phonetic form – surface form pairings), it seems that values between -2 and 4 s/b are preferred for both vowel shapes and that sporadic lower and higher values for initial pitch are preferred for one or the other vowel shape. We know that most of our data occurs between -2 and 4 s/b, and this is made clear in Fig. 6.2, which replicates Fig. 6.1 except that, in this new graph, the size of the symbol is directly correlated with how many tokens in the data set had that phonetic value for that surface form. No symbol means no data tokens.

Figure 6.2: Mean ranking values of cue constraints that regulate initial pitch, where size of label denotes number of tokens
GLOTTALIZED = ‘g’ and black; HIGH TONE = ‘h’ and gray



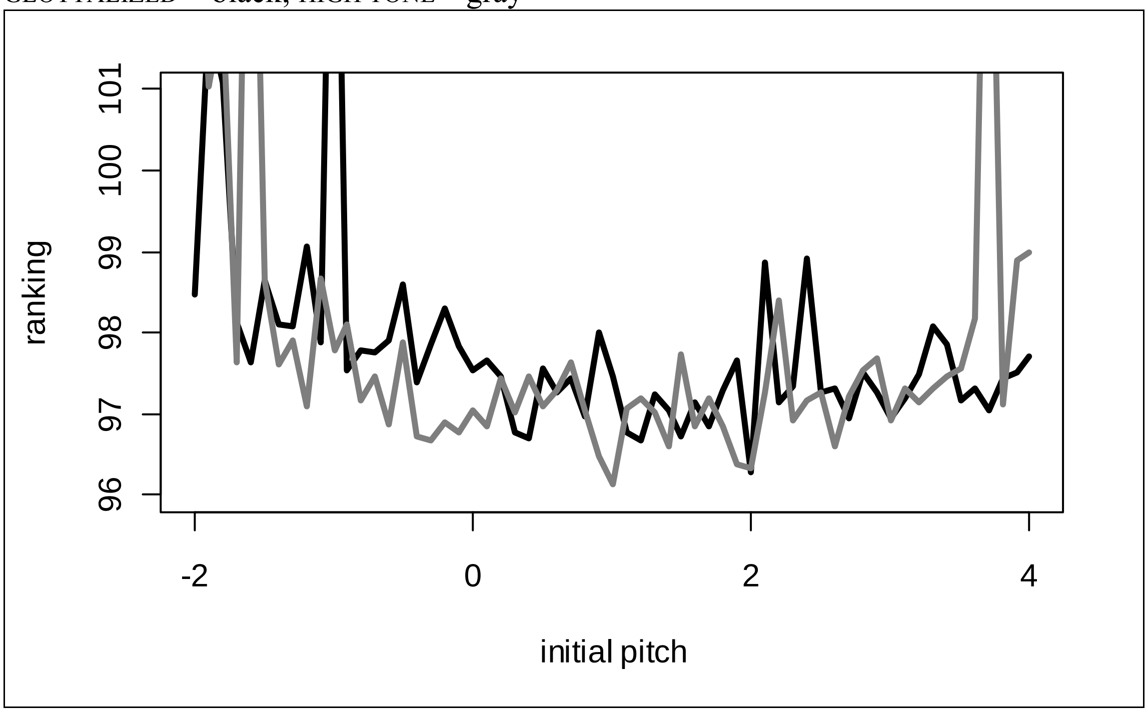
In addition to showing us the range where most of the data occurs, this graph also shows us what is causing the erratic peaks and troughs of the ranking values at the extreme ends of the initial pitch scale. We see that, wherever there is a sporadic dip in ranking value, there is some token where that particular surface form is produced with that particular initial pitch value. The reason the dips are so irregular is because of the sparseness of the data. For example, there appears to be one and only one token of a HIGH TONE vowel being produced with an initial pitch value below -6 s/b. This one token thus causes a low mean ranking value for the constraint that penalizes the pairing of that specific initial pitch value with /hi/, but since there are no other tokens with initial pitch values in that area, the other cue constraints retain higher ranking values.

Another interesting pattern exemplified in the figure above is that, whenever there is a sporadic dip in the ranking value for a cue constraint that refers to /gl/, the constraint that penalizes the pairing of the same initial pitch value with /hi/ will have an extremely high ranking value (and vice versa). This pattern is caused by learning through perception tableaus, as discussed in §5.3. If the data set only contains one token with an initial pitch of -6.2 s/b and that is a token of a GLOTTALIZED vowel, then 100% of the tokens with -6.2 s/b are GLOTTALIZED vowels. This means that if the perception grammar is tested with an initial pitch of -6.2 s/b and the grammar predicts a HIGH TONE vowel to win, a ranking adjustment will occur that will increase the ranking value of */hi/,[-6.2] and decrease the ranking value of */gl/,[-6.2]. The opposite ranking adjustment will never happen.

It is thus clear that, with such a small data set, the inconsistencies of the data will correlate with inconsistencies in ranking values. This means that we need larger data sets where these inconsistencies are evened out in order to define meaningful grammars with constraints that refer to small categories of phonetic dimensions. If we focus on the area where we do have a lot of data – the initial pitch values between -2 and 4 s/b – we will see that some of the patterns we expect to see have started to emerge. Fig. 6.3 zooms in on the ranking values of just those cue constraints that reference initial pitch values between -2 and 4 s/b. In this figure, there are still sporadic peaks and troughs, but they are less dramatic. Recall that, when using the wider categories, [2] was the preferred initial pitch value for both HIGH TONE and GLOTTALIZED vowels. With those categories, [2] referred to those values greater than [0] and less than or equal to [2]. If we look at the ranking values for the constraints that refer to initial pitch values between [0] and [2], we see that they tend to have low ranking values for both surface forms. However, the range between -2 and 0 shows that these values are more preferably paired with HIGH TONE vowels, and this is what we want.

Speakers are more likely to produce HIGH TONE values with these initial pitch values, and listeners are more likely to map such an initial pitch value onto a HIGH TONE vowel. We also expect to see initial pitch values higher than [2] to be more preferably paired with GLOTTALIZED vowels. This pattern has not emerged as clearly, though most of the pitch values around [4] are more preferably paired with GLOTTALIZED vowels.

Figure 6.3: Mean ranking values of cue constraints that regulate initial pitch in the range from -2 to 4 s/b
GLOTTALIZED = black; HIGH TONE = gray



The above discussion has pointed to the limitations of a small data set. We simply cannot use a data set of 685 tokens for the analysis of the phonetics-phonology interface with constraints that are as finely grained as this interface requires. However, I hope the above discussion has shown how promising work with larger data sets is. If some of the desired patterns emerge even with a data set of 685 tokens, we should expect the accuracy of our analysis with finely grained constraints to increase as the data set increases. Because the acquisition of such data takes considerable time and money, I hope this discussion has given reasons to expect such research to be fruitful.

6.1.2 Evaluation of Multiple Dimensions vs. Individual Dimensions in Production

In §5.6.1 I made the claim that it did not matter if we looked at all of the phonetic dimensions simultaneously or at each one individually when evaluating production tableaus. In this section I will provide the evidence for that claim. For phonologists interested in exploring the phonetics-phonology interface, I think this is good news. It means that we can do insightful analysis of production by focusing on only the dimensions of interest without worrying that the inclusion of other phonetic dimensions would lead to a different constraint ranking.

Let's consider a simplified grammar with three binary phonetic dimensions: [±glottalized], [±high pitch], [±large pitch span], which will be abbreviated [±g], [±h], and [±l]. Cue constraints penalize the pairing of each of each phonetic value ([+g], [–g], [+h], [–h], [+l], [–l]) with each possible surface form (/gl/ and /hi/). In (6.1) we see a point of evaluation for our hypothetical grammar, with the input /gl/. The winning candidate is [+g –h +l], and as this tableau illustrates, the winning candidate of each tableau is always the candidate that has the optimal value for each of the phonetic dimensions involved. Because */gl/,[-g] dominates */gl/,[+g], the winning candidate must be [+glottalized]. Similarly, */gl/,[+h] » */gl/,[-h] and */gl/,[–l] » */gl/,[+l], and so the winning candidate is [-high pitch] and [+large pitch span].

(6.1) hypothetical point of evaluation

	/gl/	*/gl/, [-g]	*/gl/, [+h]	*/gl/, [-l]	*/gl/, [-h]	*/gl/, [+l]	*/gl/, [+g]
a.	[+g +h +l]		*!			*	*
b.	[+g +h –l]		*!	*			*
c. ☞	[+g –h +l]				*	*	*
d.	[+g –h –l]			*!	*		*
e.	[–g +h +l]	*!	*			*	
f.	[–g +h –l]	*!	*	*			
g.	[–g –h +l]	*!			*	*	
h.	[–g –h –l]	*!		*	*		

Now, consider this tableau in the context of a learning step with the GLA. If the current learning datum is not [+g –h +l], the interim grammar has made a wrong prediction and a ranking adjustment will occur. For example, if the learning datum is [+g +h +l], the learner will compare line (a), the learning datum, to line (c), the (incorrect) output of the grammar and adjust all and only those constraints that favor either the learning datum or the output of the grammar, as shown in (6.2). Here we see that the only constraints that can decide between these two candidates are those constraints that regulate the only phonetic dimension that differs between the two candidates: [±h].

(6.2) learning step for /gl/ → [+g +h +l]

	/gl/	*/gl/,[-g]	→ */gl/,[+h]	*/gl/,[-l]	← */gl/,[-h]	*/gl/,[+l]	*/gl/,[+g]
a. ✋	[+g +h +l]		*!			*	*
c. ☞	[+g –h +l]				*	*	*

On the other hand, if [–g +h –l] (line f) had been the learning datum, the ranking values of all the constraints of our constraint set would be adjusted, because the output of the grammar and the learning datum would differ by each phonetic dimension, as shown below.

(6.3) learning step for /gl/ → [–g +h –l]

	/gl/	→ */gl/,[-g]	→ */gl/,[+h]	→ */gl/,[-l]	← */gl/,[-h]	← */gl/,[+l]	← */gl/,[+g]
c. ☞	[+g –h +l]				*	*	*
f. ✋	[–g +h –l]	*!	*	*			

We can make the generalization that, when a ranking adjustment occurs, all and only those constraints that regulate the dimensions that differ between the learning datum and the output of the grammar will be adjusted. This means that if we considered each phonetic dimension separately, we would get the same ranking adjustments. In (6.4) below, we see three different tableaus, one for each phonetic dimension. The point of evaluation is the same as the one illustrated in (6.2), and so the optimal candidate is [+g –h +l]. If the learning datum is [+g +h +l], which differs from the optimal candidate only by [±h], we see that a ranking adjustment only occurs for */gl/,[+h] and */gl/,[-h]. This is because the learning datum is the same as the optimal candidate for the other phonetic dimensions.

(6.4) learning step for /gl/ → [+g +h +l], each phonetic dimension evaluation individually

/gl/	*/gl/, [-g]	*/gl/, [+g]
✋/☞ [+g]		*
[–g]	*!	

/gl/	→ */gl/, [+h]	← */gl/, [-h]
✋ [+h]	*!	
☞ [–h]		*

/gl/	*/gl/, [-l]	*/gl/, [+l]
✋/☞ [+l]		*
[–l]	*!	

Similarly, if the learning datum is [–g +h –l], which differs from the optimal candidate by each phonetic dimension, the ranking values of all the cue constraints will be adjusted, whether we look at each phonetic dimension simultaneously, as in (6.5), or separately, as below:

(6.5) learning step for /gl/ → [–g +h –l], each phonetic dimension evaluation individually

/gl/	→ */gl/, [-g]	← */gl/, [+g]
☞ [+g]		*
✋ [–g]	*!	

/gl/	→ */gl/, [+h]	← */gl/, [-h]
✋ [+h]	*!	
☞ [–h]		*

/gl/	→ */gl/, [-l]	← */gl/, [+l]
☞ [+l]		*
✋ [–l]	*!	

If we add articulatory constraints to the model, this does not change the generalizations just made. Each articulatory constraint will interact with all and only those cue constraints that regulate the same phonetic dimension. For example, the constraint *[glottal stop] will interact with the cue constraints that regulate glottalization, but it will not effect the cue constraints that regulate initial pitch. Thus, we can use articulatory constraints and the same ranking values will be developed whether we look at each phonetic dimension separately or at once.

When constraints refer to multiple phonetic dimensions, they can affect the ranking values of other constraints that refer to any of phonetic dimensions mentioned by the interactive constraint. For example, in the analyses of Chapter 5, I used articulatory constraints that penalized the pairing of creaky voice with small pitch spans. Such constraints can affect the ranking values of other constraints that reference either glottalization or pitch span.

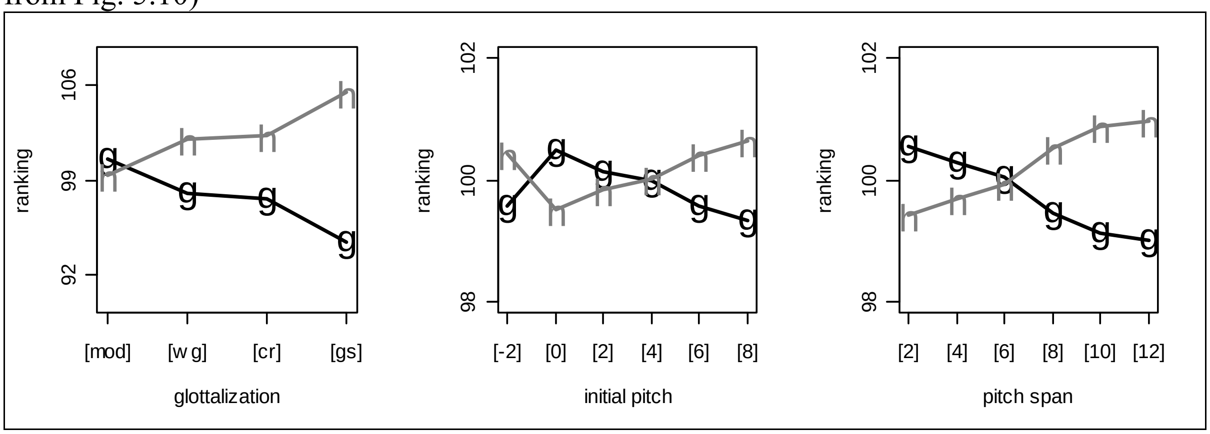
To summarize, we can do analysis with the production grammar by referring to only one phonetic dimension at a time as long as we do not think there are constraints that regulate the interaction of that dimension with some other dimension. This means that the phonologist who is interested in, say, vowel length, can analyze the production of just vowel length without worrying about ranking values of constraints that regulate, say, nasalization. This allows us to focus our attention on individual phenomena of interest and still obtain meaningful analyses.

6.1.3 Evaluation of Multiple Dimensions vs. Individual Dimensions in Perception

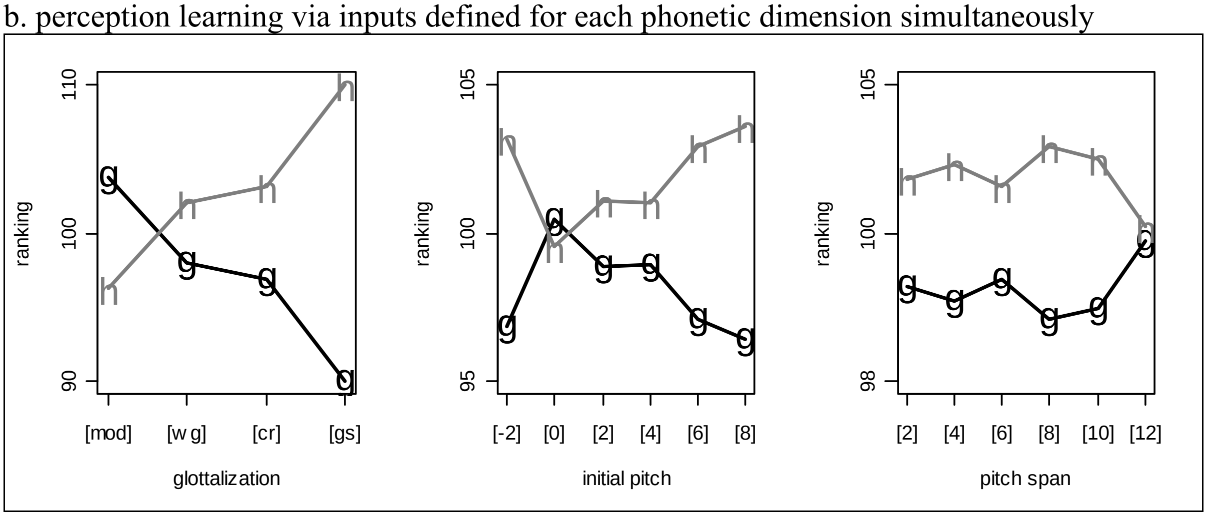
We have just seen how the evaluation of multiple dimensions in production leads to the same results as the evaluation of each dimension individually. As was explained in §5.4.1, this is not so with perception tableaus. We thus have to decide which method is more appropriate. In this section I present the results of a learning simulation with perception learning where the inputs to the perception tableaus refer to each phonetic dimension at once and compare those results to what we saw in §5.4 as the result of perception learning where the inputs are defined by one phonetic dimension at a time.

In Fig. 6.4a, we see the mean ranking values of the cue constraints that were developed through perception learning in §5.4. In this learning simulation, the input to a perception tableau was defined by one and only one phonetic dimension, either glottalization, initial pitch or pitch span. However, if we define inputs with all phonetic dimensions simultaneously and run the same type of learning simulation, we get the ranking values shown in Fig. 6.4b.

Figure 6.4: Ranking values developed by two different types of perception learning
a. perception learning via inputs defined for each phonetic dimension individually (repeated from Fig. 5.10)



b. perception learning via inputs defined for each phonetic dimension simultaneously



The surprising result that is represented in Fig. 6.4b is that almost all initial pitch and pitch span values are preferably mapped onto GLOTTALIZED vowels in perception. This is of course not how real listeners use the cues of initial pitch and pitch span. The cause of these low ranking values for GLOTTALIZED vowels and high ranking values for HIGH TONE vowels is that all constraints can potentially undergo a ranking adjustment even if only one phonetic dimension causes the grammar to make an incorrect prediction. For example, consider a point of evaluation where the input contains a value for glottalization that causes the surface form /gl/ to incur a fatal violation. In (6.6) below, the learner has heard a production of /gl/ with [modal voice]. The presence of modal voice in the input rules out the mapping of this input onto a GLOTTALIZED vowel, which means the grammar has made a wrong prediction. This error results in a ranking adjustment for each constraint. Notice that the ranking values of all of the constraints that refer to /hi/ are increased, while the ranking values of all the constraints that refer to /gl/ are decreased. If this scenario keeps happening, the rankings values of the constraints in the '/hi/ set' are perpetually increased while those in the '/gl/ set' are perpetually decreased, which is what led to the patterns represented in Fig. 6.5.

(6.6)

	→	←	←	→	→	←
[modal voice, in =4, sp = 6]	*/gl/, [mod]	*/hi/, [in=4]	*/hi/, [sp=6]	*/gl/, [in=4]	*/gl/, [sp=6]	*/hi/, [mod]
☞ /hi/		*	*			*
✋ /gl/	*!			*	*	

This result is (at least in part) a consequence of using completely arbitrary surface (phonological) forms. Because the forms /hi/ and /gl/ do not specify any phonological detail, each possible phonetic value is related to each of these surface forms with a different cue constraint. This would not necessarily be the case if these phonological forms were not completely arbitrary, i.e. if the surface forms of these two vowel shapes were /v́v/ (/hi/) and

/v́v̰/ (/gl/) as proposed in Table 2.21. In this case, it would seem more appropriate to use cue constraints that penalize the pairing of various initial pitch values with the phonological form /v́/ (and not with the two arbitrary forms /gl/ and /hi/).

In (6.7) we see a perception tableau where tone, glottalization, and length are specified with phonological markers. Cue constraints relate these phonological markers to the relevant phonetic values. In this tableau the same input is used as in (6.6), except that pitch span is ignored for simplicity. This time the mismatch between the winning candidate (/hi/, /v́v/) and the learning datum (/gl/, /v́v̰/) causes a ranking adjustment to occur only with the cue constraints that regulate glottalization. Because the two phonological forms have the same tonal marker, the constraints that regulate initial pitch cannot decide between the candidates.

(6.7)

	→		←
[modal voice, in = 4]	*/v̰/, [mod]	*/v́/, [in=4]	*/v/, [mod]
☞ /v́v/		*	*
✋ /v́v̰/	*!	*	

It is thus clear that there is no difference between the evaluation of multiple phonetic dimensions at once and the evaluation of each phonetic dimension individually when perception tableaus use candidates that are not completely arbitrary but are instead specified for various phonological features. However, the use of such specified phonological inputs would have complications for Yucatec Maya because, even though it is likely that both GLOTTALIZED and HIGH TONE vowels are marked for high tone, the production data shows that the two vowel shapes are not produced with same distribution of pitch values.

The question for future research is whether or not, when using arbitrary surface forms, the simultaneous evaluation of all phonetic dimensions in the input leads to desirable constraint rankings. In the case of Yucatec Maya, I do not think that it does. The constraint rankings in Fig. 6.4 capture none of the generalizations that we were able to make about the data set and thus do not seem to be an accurate portrayal of what the Yucatec Maya language learner would do.

6.1.4 Weighted Constraints

It was shown in Chapter 5 that the data from Yucatec Maya could be accounted for Boersma's Bidirectional Model of the phonetics-phonology interface. Specifically, the types of constraints proposed by Boersma could be ranked in such a way as to accurately predict the results of both production and perception tasks, even though the learning strategy proposed by Boersma had to be altered in order to reach an accurate adult production grammar. In this section I discuss how Boersma's model could be adapted to use weighted (instead of ranked) constraints and the advantages and disadvantageous of such an adaptation. It is my conclusion that parts of the Yucatec Maya data cannot be accounted for with weighted constraints.

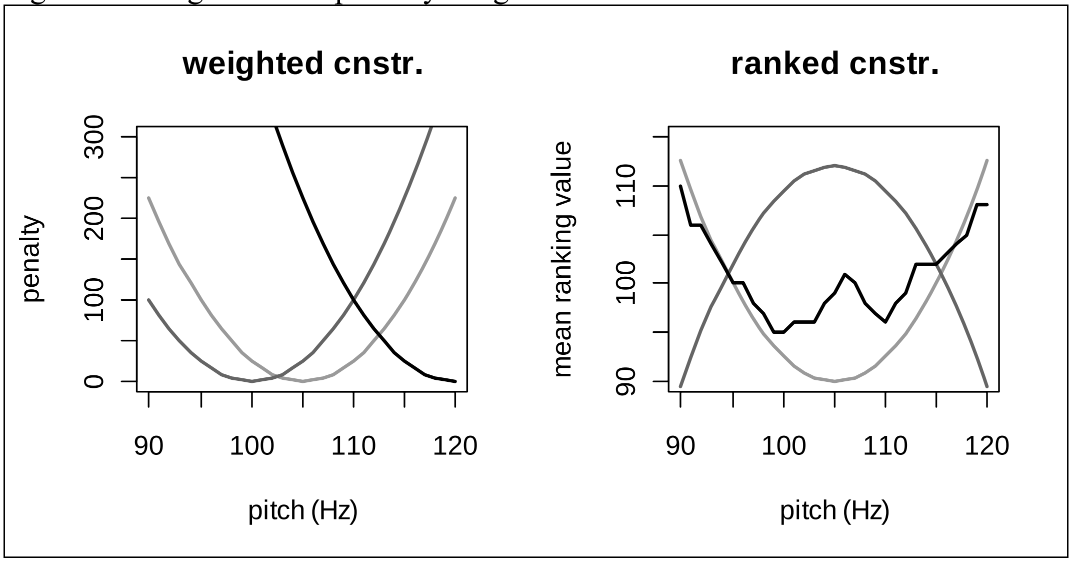
With regard to the cue constraints, bidirectional constraints that relate a phonological form to a phonetic form are advantageous for the analysis of pitch and glottalization in Yucatec Maya. This is because the same phonetic cues that are correlated with vowel shape in production are also used by the listener in perception. It is thus clear that a model of the phonetics-phonology interface should include bidirectional constraints, but we could conceive of other types of bidirectional constraints than Boersma's ranked */y/,[x] cue constraints.

In §4.2.2 we looked at Flemming's (2001) Unified Model of phonetics and phonology. In this two-level model, the underlying form is mapped directly onto a continuous phonetic form (in production) through the use of weighted constraints. An example of one of Flemming's constraints is IDENT($F2_C$) which says that consonant should be produced with an F_2 value that is equal to its target F_2 value. This constraint is not ranked with respect to other constraints but instead it assigns violation marks based on its weight (*w*), following the equation $w(F2_C - F2_{target})^2$. While Flemming's model was not originally designed to be a model of perception, his constraints could certainly work in a bidirectional fashion. For some input F_2 value, the cost of mapping it onto some consonant C could be determined by the same equation with the same weight: $w(F2_C - F2_{target})^2$, where $F2_C$ is the input value and $F2_{target}$ is the known target for that consonant.

The immediate advantage of such a system is that we drastically reduce the number of constraints, making the analysis simpler and easier to understand. For example, instead of all of the cue constraints that penalize pairings of each initial pitch value with each surface form we would only need one constraint: IDENT($F0_V$), which would assign violation marks following $w(F0_V - F0_{target})^2$. However, there are many questions that need to be answered about these types of constraints before any analysis can be pursued with such a system. First, is there only one IDENT($F0_V$) (or IDENT($F2_C$)) constraint? If so, this implies that the penalty for deviating from the target is the same regardless of the input. If not, is there a universal set of them or are there some language-specific details that go into the constraint? If we consider the data from Yucatec Maya, we saw that the initial pitch values of GLOTTALIZED vowels were more variable than those of HIGH TONE vowels because of the effect of creaky voice on pitch. Does this mean that the penalty for F_0 deviations is not as strong for GLOTTALIZED vowels (and hence that there are multiple IDENT($F0_V$) constraints), or can we motivate the deviations with other constraints, such as those that correlate F_0 and creaky voice (meaning that there are not multiple IDENT($F0_V$) constraints)? Additionally, how does the learner determine what the target value is (for F2 or F0 or some other phonetic dimension)? Where is this knowledge stored and does/can it change over time? Would stochastic evaluation of weights (or target values) allow for the same patterns as stochastic evaluation of constraint rankings?

I will not attempt to answer these questions at this time, but I believe that in the meantime we can discuss the advantages and disadvantages of each constraint type, both in general terms and as related to the Yucatec Maya data. First, in general terms, it is clear that Boersma's cue constraints (henceforth "ranked constraints") can account for more types of data than Flemming-style weighted constraints (henceforth "weighted constraints"). If we consider just one phonetic dimension, say pitch, one weighted constraint (i.e. IDENTF0) specifies a target and penalizes deviations from that target, with further deviations receiving larger penalties. Some examples of how this constraint would assign penalties to different phonetic forms are shown in the left-hand graphs of Fig. 6.5. Here we see three different versions of IDENTF0: one that specifies a target of 100 Hz, one that species a target of 105 Hz, and one that specifies a target of 120 Hz. The lines of this graph show the penalty associated with each pitch value when the weight of each constraint is 1. For example, if the target value is 100 Hz, a phonetic form with 90 Hz receives a penalty of 100, as determined by the equation $(90\text{-}100)^2$.

Figure 6.5: Regulation of pitch by weighted vs. ranked constraints



A set of ranked constraints is needed to regulate the same phonetic dimension. Because each member of this set is freely rankable with respect to the other members, the ranking of the set of constraints could mimic the job of one weighted constraint, i.e. the set could be ranked in such a way that one target value is preferred and deviations from this target are more disprefered the larger the deviation, as shown by the light gray line in the right-hand graph in Fig. 6.5. In the right-hand graph, the light gray line shows how a set of Boersma-style ranked cue constraints could mimic the pattern predicted by one Flemming-style weighted constraint (as denoted by the light gray line in the left-hand graph). But this is not the only way for cue constraints to be ranked. A set of ranked cue constraints could follow the pattern denoted by the dark gray line, where peripheral values are preferred and central values disprefered. Additionally, a set of ranked cue constraints could follow the pattern denoted by the black line, where value near 100 or 110 are the most preferred, and where deviations from these values are more disprefered, but in a less consistent manner.

It is thus clear that nature of Flemming-style weighted constraints is such that they express a desire to reach some target and penalize deviations from that target in a way that is relative to the size of the deviation. Boersma-style ranked constraints do no reference any target. The overall ranking of the constraints may indicate a target and may indicate a stronger dispreference for values that are further from that target than those that are closer to the target (such as the ranking of cue constraints that regulate initial pitch in Yucatec Maya, see Fig. 5.13), but they need not indicate a single target nor progressively penalize deviations from some target (as with the ranking of cue constraints that regulate pitch span in Yucatec Maya, see Fig. 5.13).

It seems that weighed constraints may undergenerate in that they cannot account for all the patterns that are attested in real languages, while ranked constraints overgenerate in that they can account for those patterns that are attested in real languages and likely more. Because the rankings of cue constraints that were deemed optimal in accounting for pitch span in the Yucatec Maya data can not be mimicked with one weighted constraint, I believe the problem of undergeneration makes this type of weighted constraint ill-suited for analysis of the phonetics-phonology interface. Furthermore, there are many ways to address the problem of overgeneration, such as the use of theory-internal devices (e.g. fixed ranking hierarchies (Prince and Smolensky (1993/2004) or target constraints (Wilson 2001) in OT) or by appealing to diachrony and the processes of language transmission from one generation to the next (Myers 2002). Myers (2002) argues that tinkering with the constraint set or the possible constraint rankings will never account for all of the gaps in factorial typology (i.e. for all instances of overgeneration) and thus that these gaps occur because natural diachronic processes will never lead to the development of such a language. This means that the burden of controlling overgeneration does not necessarily fall on the synchronic grammatical devices of a particular theory.

One final point is relevant with regard to articulatory constraints. Consider a Flemming-style MINIMIZEEFFORT constraint of the form $F0_{max} = F0_{min}$, which is a constraint that would penalize any pitch span above 0, with larger pitch spans receiving larger penalties. Such a constraint would not be able to account for the anti-Paninian rankings that were developed for the set of *[large pitch span] articulatory constraints (see §5.7), because this weighted constraint specifies [pitch span = 0] as the target and progressively penalizes deviations from this target. According to the mean ranking values of the articulatory constraints in Table 5.17, a pitch span of [12] is more preferable than smaller pitch spans, but a pitch span of [4] is more preferable than [6] (which is more preferable than [8], which is more preferable than [10]). There is no way for a Flemming-style MINEFF constraint to account both the fact that [sp = 4] is better than [sp = 6] and the fact that [sp = 12] is better than [sp = 4] because there is no one target value that allows for this pattern. Again, the use of weighted constraints leads to a model that undergenerates. It would be beneficial for future research to explore the consequences of constraint ranking versus constraint weighting in the context of the Bidirectional Model, but I believe this section has provided evidence that freely rankable cue constraints can account for language patterns (some of which are found in the Yucatec Maya data) that weighted constraints cannot.

6.2 Predictions of the Model

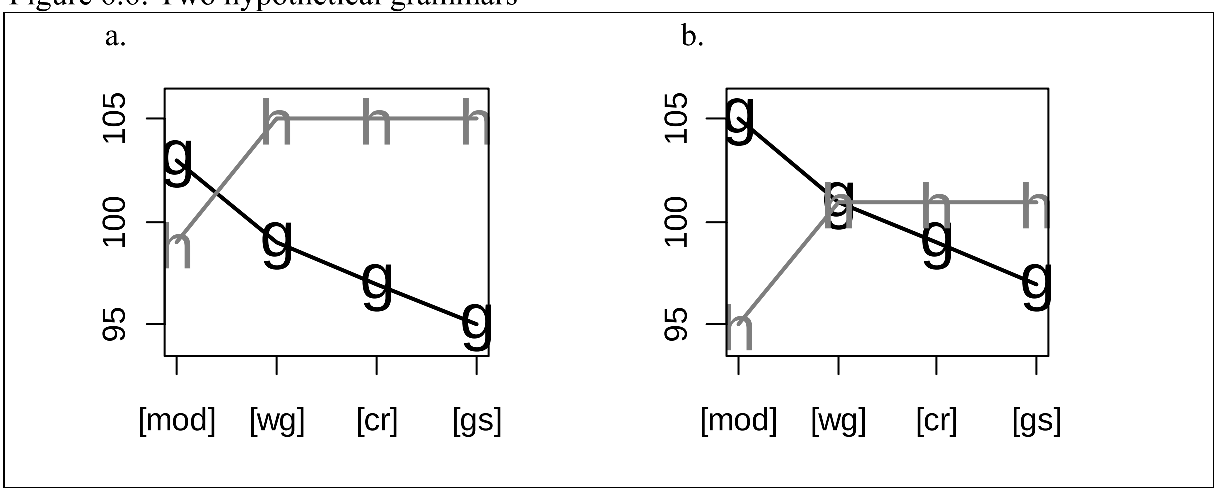
6.2.1 Synchronic

We know that symmetries often occur between production and perception. The Bidirectional Model elegantly accounts for these symmetries. But just how symmetrical are production and perception? The addition of production learning to the model predicts that there is not a one-to-one correlation between possible production and perception grammars.

In other words, production does not simply fall out of perception. This means that production and perception can (to a certain degree) vary independently.

If we consider again the two hypothetical mini-grammars that were presented in §5.2.3 (repeated in Fig. 6.6), we see that these are two identical production grammars but two very different perception grammars. Both production grammars predict for HIGH TONE vowels to be commonly mapped onto modal voice and for GLOTTALIZED vowels to be commonly mapped onto some form of glottalization. In perception grammar (a), modal voice will likely be mapped onto a HIGH TONE vowel while all forms of glottalization will be likely mapped onto a GLOTTALIZED vowel, however, in perception grammar (b), the glottalization cues are not as robust – weak glottalization will be associated with either vowel shape half the time and creaky voice will not consistently be mapped onto a GLOTTALIZED vowel.

Figure 6.6: Two hypothetical grammars



A question for future research is how much production and perception can independently vary. If we consider the production grammars above, we can see that there would be no way to keep the same production grammars (as defined by the relations among the constraints on a given line) and move the lines up or down to make glottalization a better cue to a HIGH TONE vowel than modal voice. As long as the HIGH TONE line has a dip at modal voice and the GLOTTALIZED line has a dip at a glottal stop, modal voice will be the best cue for perceiving a HIGH TONE vowel and a glottal stop will be the best cue for perceiving a GLOTTALIZED vowel. Just how good these cues are, though, is a function of how close together the various parts of the lines are.

In this way, the use of bidirectional cue constraints predicts for symmetries to occur, but the simultaneous learning from production and perception tableaus predicts that these symmetries do not have to be exact. There can be multiple perception grammars that equate with the same production grammar and also multiple production grammars that equate with the same perception grammar.

6.2.2 Diachronic

It is currently well-accepted that sound change can result from mistakes and/or variation in the signal; noisy environments lead to errors of transmission, and over generations, if these errors pile up, they are no longer errors. Earlier discussions on this topic centered on the idea of *phonologization* (e.g. Hyman 1977, Ohala 1981, 1993), where intrinsic variation in the signal is reinterpreted as extrinsic and hence becomes a controlled property of speech in the new generation that performed the phonetic reanalysis. An oft-mentioned example is how tonal contrasts arise from intrinsic variation in pitch due to adjacent consonants. This variation is at first noncontrastive and a simply a product of coarticulation, but it is reinterpreted as an intentional property of the vowel, thus denoting a phonemic tone contrast.

A detailed approach to the noisy channel problem comes from Blevins (2004, see also Blevins and Garrett 1998), who focuses only on those situations where the listener makes an error. She categorizes three types of sound change that result from misperception: *change* results from misperception due to phonetic similarity, *chance* occurs when the phonetic signal is phonologically ambiguous and the learner interprets a different phonological structure than the speaker intended, and *choice* refers to when a language learner selects a different prototype from a list of (correctly perceived) variant forms.

It is thus clear that, even though there are other forces involved in creating and controlling language change, sound change can result from either misspeaking or mishearing. When using a model with stochastic evaluation, we have to reevaluate what we mean by *misspeaking* and *mishearing*. Consider the situation where stochastic evaluation leads to an unusual ranking and hence to the selection of an unusual candidate as optimal, which could happen in production or perception. In this case, a surface form – phonetic form pairing was selected by the grammar even if it is not normally an ideal surface form – phonetic form pairing. We could argue that such situations are not errors because they are controlled by the grammar, but I will use the term error to refer to any surface form – phonetic form pairing that is expected to occur less than some arbitrarily small percentage of the time.[69] I will call this type of error a *competence error*. The other type of error occurs when the speaker or listener messes up in some way that is not determined by the grammar, i.e. if the listener makes a *performance error*. Such an example might be a case in which the speaker produces an aspirated stop but the listener simply does not hear the aspiration (because it was not long enough or because there was some other sound in the background, etc.) and thus does not

[69] For example, when PRAAT is used to calculate the predict output distributions for some grammar, it runs 100,000 simulations and then shows how often each input – output pairing occurred out of 100,000 times. As we saw in §5.6.3, there were outputs that were predicted to occur 0 out of 100,000 times for that production grammar. This does not mean that these outputs will never be deemed optimal by that grammar, just that they are expected to occur less than .001% of time. I am thus proposing that we should think of such outputs as errors even when their production is dictated by the grammar.

start with a phonetic form with aspiration when performing the task of comprehension.

When we think about errors in this way, it is unclear which type of error the *change* category is. Blevins (2004: 32) provides an example where the speaker says [anpa] but the listener hears [ampa]. In the context of the Bidirectional Model and the phonetics-phonology interface, we might question what it means to "hear [ampa]". Does the listener start with [ampa] as the auditory form that is the input to the first stage of perception? Or, does the listener start with [anpa] and (incorrectly) map it onto the surface form /ampa/? Is the misperception in the head (a performance error) or in the grammar (a competence error)?

I believe Blevins intends for *change* to be a performance error. However, given stochastic evaluation which leads to the idea of competence errors, I would like to consider the implications of competence errors for language change. Furthermore, I wish to focus on those competence errors made by the speaker, which is a different approach than both phonologization and evolutionary phonology, where misperception is the catalyst of sound change.

Competence errors are more likely to occur when conflicting constraints have closer mean ranking values. Thus, the competence errors that the adult speaker is more likely to produce in the presence of a child language learner are those errors that are caused by a reverse in the ranking order of two constraints that are usually ranked one way but have close enough mean ranking values that a reverse ranking can happen (some arbitrarily small percentage of the time). Depending on where the language learner is in the acquisition process, we might expect for the child's current grammar to also have the two constraints that caused the error with somewhat close mean ranking values. If the learner takes a learning step on the basis of the adult's competence error, the learner's grammar will be adjusted to look more like the point of evaluation that caused the error.

Let's consider a simple hypothetical example. In some language, the input /bid/ tends to be mapped onto [bid], but some speaker says /bid/ as [bit] (due to an aberrant ranking of */d/,[d] » */d/,[t]) in the presence of a language-learner.[70] This learner then tests the interim grammar with both a production and perception tableau, as shown in (6.7), and ranking adjustments are made. What is interesting is that, because the learner checks both production and perception, the constraint */d/,[t] is demoted twice. Because the error was a competence error, it arose because the constraints */d/,[d] and */d/,[t] were already ranked relatively closely together, and now the learner has demoted */d/,[d] twice, putting it even closer to */d/,[t] (which was also promoted once).

(6.7) an adult's competence error: /bid/ → [bit]

a. production tableau

/bid/	→ */d/,[t]	← */d/,[d]
☞ [bid]		*
✋ [bit]	*!	

b. perception tableau

[bit]	→ */d/,[t]	← */t/,[t]
☞ /bit/		*
✋ /bid/	*!	

Thus, the simultaneous learning from production and perception tableaus predicts that a constraint's ranking value can be adjusted twice on the basis of one learning datum. For performance errors, which may result in pushing constraints closer together that are already quite far apart, this is not that much different from adjusting a ranking once on the basis of

[70] In this hypothetical example, I use phonetic forms that look like phonological forms for simplicity. The form [bit], for example, should be taken to mean *the phonetic form with acoustic cues that are normally associated with /bit/*.

one learning datum. For competence errors, on the hand, which result because of how close the ranking values of two constraints are, the prediction here is that such errors can affect the grammar at an accelerated rate. The learner can more rapidly learn to repeat competence errors because of the close ranking between the two constraints that caused the error in the first place.

This section has pointed to two ideas that are worth exploring in the field of historical linguistics. The first is that there is such a thing as a competence error. Stochastic evaluation predicts that both speakers and hears will, every once in a while, produce or hear things that are unexpected, and they will do this *because the grammar told them to* and not because of some performance error. The second is that the simultaneous learning via production and perception tableaus predicts that those competence errors that are more likely to occur can be learned relatively fast.

7

CONCLUSIONS

7.1 Phonetic Description of Yucatec Maya

Chapters 2 and 3 document a variety of facts about the production and perception of vowel shape in Yucatec Maya. This is an area that had not been thoroughly explored in the phonetics literature, and so I will present the conclusions that we can draw from these production and perception studies here.

First, it is clear that vowel shape varies by dialect. We can identify at least two distinct dialect groups that I have referred to as the western dialect and the eastern dialect. We should be cautious about these labels because I do not yet have data from speakers from the states of Quintana Roo (in the east) or Campeche (in the southwest). The western dialect is defined by the data from Mérida and Santa Elena, while the eastern dialect is defined by the data from Sisbicchén, Xocén, and Yax Che. I will first discuss the generalization that we can make about the western dialect in §7.1.1 and then I will address the eastern dialect in §7.1.2. I then summarize the findings with respect to idiolectal variation in §7.1.3 and with respect to laryngeal complexity in §7.1.4.

7.1.1 Vowel Shape in the Western Dialect

The production of the features of length and pitch by speakers from the western dialect closely resembles the descriptions in the phonological literature (as reviewed in §1.2.4). Long vowels (HIGH TONE, LOW TONE, and GLOTTALIZED) are about twice as long as the SHORT vowel. LOW TONE vowels are produced with fairly steady low pitch, SHORT vowels are produced with fairly steady mid or mid-high pitch, GLOTTALIZED vowels are produced with initial high pitch and a falling contour, and HIGH TONE vowels are also produced with initial high pitch and a falling contour. GLOTTALIZED vowels tend to have higher initial pitch and a larger pitch span than HIGH TONE vowels. The position of the vowel with respect to the phrase influences the pitch contour of HIGH TONE vowels only: in non-phrase-final position these vowels tend to have a rising contour instead of a falling one.

The production of glottalization with the GLOTTALIZED vowel is more variable than suggested by previous phonological literature. It had been assumed that the phonological representation of the GLOTTALIZED vowel is /v́ʔv/, but my production studies (and those of Avelino et al. 2007) show that a glottal stop is only very rarely produced. Instead, this vowel is more commonly produced with creaky voice or no glottalization at all (i.e. modal voice is produced throughout the vowel duration). I have thus proposed that a more accurate phonological representation for this vowel shape is /v́v̰/, and in §3.5 I discussed how this representation reflects the idea that the GLOTTALIZED vowel and the HIGH TONE vowel (/v́v/) are minimally contrastive on the basis of glottalization.

Because of the high degree of overlap in the phonetic cues that are produced for HIGH TONE and GLOTTALIZED vowels, a perception experiment was conducted in order to determine how successful listeners are at distinguishing these vowel shapes and what cues they use to do so. This experiment was discussed in Chapter 3. The results indicate that, when listening to natural stimuli, the listeners can do better than chance at discriminating between the two vowel shapes and that they use all the cues of glottalization, initial pitch, and pitch span to make their choice. Thus, there are symmetries between production and perception: the phonetic cues that distinguish the vowel shapes in production are the same cues that the listener attends to in perception. The consequences of this for the grammar of Yucatec Maya are summarized in §7.2.

7.1.2 Vowel Shape in the Eastern Dialect

Participants from the eastern dialect did not produce vowel shape as it is described in the phonological literature. With respect to glottalization, these participants behaved like participants from the west in that the GLOTTALIZED vowel is most commonly produced with creaky voice or no glottalization and only rarely with a glottal stop. In terms of length and pitch, though, there are many differences that distinguish this dialect group from the west. In this section I will generalize over all speakers from the east, even though in Chapter 2 we saw some cases where the east did not behave as a homogenous group.

The long vowels were all longer than the SHORT vowel in the east, but not all long vowels were the same length. The HIGH TONE vowel was consistently the longest vowel shape. For example, in production study 2 (frame sentences), the average vowel length of HIGH TONE vowels in phrase-final position was 224 ms, while LOW TONE and GLOTTALIZED vowels were on average 178 and 190 ms, respectively (and SHORT vowels averaged 112 ms). The question is then whether or not a three-way length contrast is utilized by the eastern dialect. While such a contrast is rare,[71] there is a reason to believe this is exactly the case for

[71] A variety of languages have been claimed to have a three-way contrast in vowel length, but there are also skeptics who claim such contrasts do no exist. A three-way vowel contrast is described for Dinka (Nilo-Saharan) in Anderson (1987), and Remijsen and Gilley (2008) present acoustic evidence that supports this analysis. Mixe (Mixe-Zoquean) is argued to utilize a three-way vowel length contrast by Hoogshagen (1959), but Wichmann (1995: §2.4) shows that there are other ways to analyze the contrast. The contrasting vowel and consonant lengths in Estonian have been long discussed. With this language, the phonetic facts are not in dispute but the question of how to represent the length contrasts is widely debated. Prince (1980) claims that the differences between long and super-long elements (consonants or vowels) are only apparent in the metrical structure and not in the segmental representations of such elements. In any case, there is evidence (though it is

these speakers. This reason comes from the results of the production study with respect to pitch.

The speakers from the east produced pitch contours for HIGH TONE vowels that were either indistinguishable from the pitch contours of the LOW TONE vowel or that had lower pitch than LOW TONE vowels. It thus seems likely that a tonal merger has occurred for these speakers. However, that does not mean that phonemic merger has occurred, because the HIGH TONE vowels are longer than the LOW TONE vowels. I am hesitant to draw the firm conclusion that a length contrast has replaced a tonal contrast in the eastern dialect because of the low numbers of participants from the east, but I believe the data that was presented in Chapter 2 is sufficient to motivate further experimentation in this area in order to figure out exactly how much dialect variation there is with respect to pitch and length in Yucatec Maya.

7.1.3 Idiolectal Variation

Another important conclusion that came out of the production studies has to do with the amount of idiolectal variation that was present. The production of glottalization was highly variable among participants; there were participants who produced GLOTTALIZED vowels with a full glottal stop half of the time and there were participants who never produced any glottalization. Furthermore, there were no consistent dialectal patterns, i.e. it is not the case that those participants who favored a glottal stop came from a different town than the participants who favored no glottalization. The production of pitch was not subject to as much inter-speaker variation, but, as shown in Fig. 2.17, there were participants from Santa Elena who produced remarkably different pitch contours than those of the majority of speakers.

---

not uncontroversial) that comes from multiple language families that suggests that three-way length contrasts are possible – but rare – with vowels.

As linguists we of course have to abstract away from idiolectal variation because our goal is to generalize about an entire language (or dialect), but we cannot completely ignore this variation. The best we can do is to take measurements from large numbers of speakers so that our data is not completely swayed by the idiosyncrasies of one individual. In the study of Yucatec Maya, this has not always been the practice. The only phonetic study done in the 20th century of pitch in Yucatec Maya (Fisher 1976) used measurements from one speaker. But one speaker cannot be relied on to accurately represent the behavior of a population. I thus wish to stress the importance of doing phonetic experimentation with as many participants as possible.

7.1.4 Glottalization and Laryngeal Complexity in Yucatec Maya

The production studies presented in Chapter 2 and the work of Avelino et al. (2007) show that the GLOTTALIZED vowel of Yucatec Maya, commonly called a 'rearticulated' vowel and represented as /v́ʔv/ or simply /vʔv/ (e.g. Bricker et al. 1998, Blair 1964, Blair and Vermont Salas 1965), is canonically produced with creaky voice. There are many Mayan languages that have a 'rearticulated' vowel phoneme, but I do not know of any phonetic study on the production of this vowel shape in other Mayan languages. It would be beneficial to obtain such data not only for the purposes of language documentation but also because such data is important for language typology and historical reconstruction.

I have claimed (in §3.5 and elsewhere) that creaky voice is the most important cue to the contrast between HIGH TONE and GLOTTALIZED vowels and that both of these vowel shapes are marked for high tone. This suggests that Yucatec Maya should be considered a 'laryngeally complex' language because tone and phonation type combine to characterize what I call vowel shape. There has been some interesting discussion in the literature about laryngeal complexity in the Otomanguean languages. In the languages of Jalapa Mazatec, Comaltepec Chinantec, and Copala Trique (Silverman 1997) and Yalálag Zapotec (Avelino 2004), vowels contrast on the basis of tone and phonation type (modal, creaky, or breathy voice). Silverman suggests that tone and non-modal phonation are always sequenced with respect to each other so that the optimal cues are available to the listener. This sequencing is true of Yucatec Maya as well – tone occurs on the first half of the vowel and creaky voice on the last half. Additionally, the results from the production studies of Yucatec Maya show that pitch values never vary during the production of creaky voice. If this is a cross-linguistic principle, it suggests that tone and creaky voice have to be sequenced because different pitch values cannot be produced during creaky voice. There are thus reasons to reexamine the Otomanguean languages in light of the data from Yucatec Maya. Furthermore, now that it is clear that Yucatec Maya is laryngeally complex, we can use this language to advance our understanding of the cross-linguistic principles of laryngeal complexity.

7.2 The Grammar of Yucatec Maya at the Phonetics-Phonology Interface

The grammar (for Santa Elena) that was developed in §5.6 was able to account for the production of pitch and glottalization with the HIGH TONE and GLOTTALIZED vowels with a high degree of accuracy. Furthermore, this grammar captures many of the intuitions about the distribution of the data that we made in Chapters 2 and 3. We saw in these chapters that the two vowel shapes significantly differ in terms of both pitch and glottalization but that glottalization was the most consistent indicator of the contrast. This is reflected in the grammar that is developed with the GLA. The mean ranking values of the cue constraints that regulate glottalization are drastically different for HIGH TONE and GLOTTALIZED vowels. The mean ranking values for the cue constraints that regulate initial pitch, on the other hand, are different for the two vowel shapes but not to the same degree. This results in a grammar that predicts for glottalization to be more consistently used by both the speaker and the listener than pitch. This is of course true of actual speakers and listeners.

In this way, the grammar of Yucatec Maya at the phonetics-phonology interface controls a variety of acoustic cues and dictates which are more important that others. This what we expect in natural speech – multiple cues integrate into a contrast, though not all cues are always present or are as tightly controlled as other cues. The fact that glottalization seems to be the most important cue to this particular contrast in Yucatec Maya is reflected in the phonological forms that I have proposed for the HIGH TONE (/v́v/) and GLOTTALIZED (/v́v̰/) vowel. Both vowel shapes are marked for high tone because both are produced with significantly higher pitch than the LOW TONE or SHORT vowels. Even though pitch also significantly differs between the HIGH TONE and GLOTTALIZED vowels, these differences are not as consistent and are not used by the listener in degraded language situations. It is for this reason that I propose that, with respect to tone, these two vowel shapes are identical at the phonological level and it is the phonetic grammar which maps them onto different distributions of pitch values. In terms of glottalization, it is clear that this feature contributes to contrast at the phonological level, and hence the GLOTTALIZED vowel is marked for creaky voice while the HIGH TONE vowel is not.

7.3 Assessment of Bidirectional Stochastic OT and the GLA

In the analyses of Chapter 5, the cue constraints and articulatory constraints (mostly) followed the formats specified by Boersma (especially 2006, 2007a). Cue constraints are defined to penalize the pairing of each possible combination of the relevant surface and phonetic forms. Articulatory constraints penalize effortful productions, however I have implemented an innovation in that I use articulatory constraints that are scalar (following de Lacy 2002, Prince 2007a-b). These types of constraints were able to be ranked in such a way that perception and production were accurately accounted for. Specifically, the bidirectional cue constraints were able to account for the symmetrical nature of pitch and glottalization in production and perception, and the articulatory constraints were able to account for the *prototype effect* (Boersma 2006), such as the fact that a [glottal stop] is not commonly produced for GLOTTALIZED vowels but is an excellent cue to the perception of GLOTTALIZED vowel.

In order to reach the most accurate ranking of the relevant constraints, the learner had to be trained on both perception and production tableaus, which is not the learning strategy proposed by Boersma (2006, see also Escudero and Boersma 2004 and Boersma and Hamann 2008). Boersma proposed that the mean ranking values of cue constraints are learned through perception learning alone, but when such a learning simulation is run with the Yucatec Maya data, the resulting grammar does not capture many of the intuitions about the production data. For example, this grammar says that an initial pitch value of [0] is very good for a HIGH TONE vowel but very bad for a GLOTTALIZED vowel; however in actual productions this initial pitch value is somewhat common for both vowel shapes. When the learner is trained on both perception and productions tableaus when learning the ranking values of cue constraints, this generalization about the data can be accounted for because both */hi/,[in=0] and */gl/,[in=0] end up with similar relatively low mean ranking values. Furthermore, language-specific production learning is necessary with the articulatory constraints, as shown by the fact that anti-Paninian rankings best account for the production data and such rankings cannot be assumed to be universal.

This is the first work in which a GLA learner is trained on actual language data in order to learn the rankings of cue and articulatory constraints in the context of Bidirectional StOT. We have thus seen that a GLA learner who is trained on production and perception tableaus can learn to mimic the distribution of phonetic forms from a real language environment and that the interaction of cue and articulatory constraints can account for these distributions. This is a positive result for the GLA and Bidirectional StOT. The number of tokens in the data set that I used limited how finely grained the constraints could be. For this reason, I was not able to account for categories of phonetic dimensions that are as small as we would like in the phonetic component of the grammar. However, the success of the model with a data set of 685 tokens indicates that, if we use larger data sets, we can expect the model to be successful with more finely grained constraints. I thus believe that the results presented here motivate continued work with the model by using real language data, especially larger data sets with larger numbers of speakers.

APPENDIX A

TARGET WORDS USED IN PRODUCTION STUDIES

A.1: Production Study 1

The following wordlist includes each Yucatec Maya target word used in production study 1 and its English and Spanish translation (the latter was printed below the Yucatec Maya word on the stimuli). All words are used in both word lists unless otherwise noted (SE = Santa Elena speakers only, nonSE = all other speakers). Nonce forms are denoted by ".." in place of English/Spanish translations; underlining denotes the CVC context from which measurements were taken in polysyllabic forms.

	i'	**a'a'**	**aa'**	**áa**
	hawk	..	..	..
	gavilán	..	..	..
	ich	**e'es**	**ook**	**óox**
	in	show	foot	three
	en	mostrar	pie	trés
	am	**a'al**	**oon**	**éem**
	spider	speak	avocado	descend
	araña	hablar	aguacate	descender
	uk'	**a'ak'**	**eek'**	**éets'**
	louse	..	star	echo
	piojo	..	estrella	eco
nonSE:	**ab**	**u'ub**	**iib**	**áab**
	..	listen	bean	..
	..	escuchar	frijol	..
SE:	**abal**	**u'ubik**	**iibil**	**áabil**
	plum	hear it	bean	grand-child
	ciruela	escucharlo	frijol	nieto
	ka'	**ti'i'**	**tsuu'**	**tsáa'**
	metate	there	aguti	rattle
	metate	allí	agutí	cascabel
	chak	**tso'ots**	**tsaap**	**cháak**
	red	hair	fuzz that causes itching	rain
	rojo	pelo	cosa áspera (que causa comezón)	llueve
	kan	**xi'im**	**tseem**	**chéel**
	four	corn	chest	rainbow
	cuatro	maíz	pecho	arco iris
	pak'	**pi'its'**	**peek'**	**púuts'**
	wall	slightly (sweet)	dog	needle
	pared	ligeramente (dulce)	perro	aguja
nonSE:	**chab**	**ta'ab**	**xiib**	**píib**
	anteater	salt	man	underground roasting pit
	oso hormiguero	sal	hombre	horno hecho bajo tierra

SE:	**chabo'**	**ta'abo'**	**xiibo'**	**píibo'**
	that anteater	that salt	that man	that underground roasting pit
	eso oso hormiguero	eso sal	eso hombre	eso horno hecho bajo tiera
	ni'	**na'a'**	**laa'**	**náa'**
	nose	..	old	..
	nariz		viejo	
	lak	**na'at**	**miis**	**máak**
	clay cup	intelligent	cat	person
	taza hecha de barro	entendimiento	gato	hombre
	nal	**mo'ol**	**maan**	**néen**
	corn	paw	buy	mirror
	elote	pata de los felinos	compra	espejo
	mak' *	**ma'ats'**	**neek'**	**láak'**
	cork	hull (corn)	seed	other
	corcho	hollejo	semilla	otro
nonSE:	**nab**	**ya'ab**	**yeeb**	**náab**
	..	a lot	fog	hand span
	..	mucho	niebla	cuarta
SE:	**nabo'**	**ya'abo'**	**yaabilaj**	**náabo'**
	..	a lot	love	that hand span
	..	mucho	amor	cuarta
	ch'o'	**k'a'a'**	**t'uu'**	**k'áa'**
	mouse	..	side (of hammock)	..
	ratón	..	lado (de hamaca)	..
	k'at	**p'u'uk**	**k'aas**	**k'áax**
	clay	jaw, cheek	ugly	forest
	barro	mejilla	feo	bosque
	k'an	**k'a'an**	**k'iin**	**k'áan**
	ripe	strong	day, sun	hammock
	maduro	fuerte	día,sol	hamaca
	ch'och'	**k'i'ik'**	**t'uut'**	**k'áak'**
	cicada	blood	parrot	fire
	cigarra	sangre	loro	fuego
nonSE:	**k'ab**	**k'a'ab**	**k'aab**	**ts'íib**
	arm	..	..	writing
	brazo	..	..	escritura
SE:	**k'abo'**	**k'a'abéet**	**k'aaba'**	**k'óoben**
	that arm	necessary	name	kitchen
	eso brazo	necesario	nombre	cocina
	k'aaba'	**ba'a'**	**baa'**	**báa'**
	name	..	..	..
	nombre	..	..	..
	bix	**ba’ax**	**beet**	**báat**
	how	what	make, do	axe
	como	que	hacer	hacha

	bin	**bu'ul**	**beel**	**bíin**
	go	bean	road	*future aspect*
	ir	frijol	camino	*el futuro*
	bak'	**bi'ik'**	**beech'**	**bóoch'**
	meat	wiggle	quail	shawl
	carne	culebrear	codorniz	rebozo
nonSE:	**bab**	**be'eb**	**baab**	**báab**
	..	a type of vine	..	swim
	..	un tipo de planta	..	nadar
SE:	**babo'**	**ba'abo'**	**baabo'**	**báabo'**
	..	..	..	..
	..	..	..	..

* Not all speakers recognized this word (*mak'*), and so for some it was a nonce form.

A.2: Production Study 2

The following wordlist includes each Yucatec Maya target word used in production study 2, followed by an English and Spanish translation of the word (the Spanish translation was printed along with the Yucatec Maya word on the stimuli).

chak	**tso'ots**	**tsaap**	**cháak**
red	hair	fuzz that causes itching	rain
rojo	pelo	cosa áspera que cause comezón	lluvia
kan	**xi'im**	**tseem**	**chéel**
four	corn	chest	rainbow
cuatro	maíz	pecho	arco iris
pak'	**pi'its'**	**peek'**	**púuts'**
wall	slightly (sweet)	dog	needle
pared	ligeramente (dulce)	perro	aguja
lak	**na'at**	**miis**	**máak**
clay cup	intelligent	cat	person
taza hecha de barro	entendiminto	gato	persona
nal	**mo'ol**	**maan**	**néen**
corn	paw	buy	mirror
elote	pata de los felinos	compra	espejo
mak'	**ma'ats'**	**neek'**	**láak'**
cork	hull (corn)	seed	other
corcho	hollejo	semilla	otro
k'at	**p'u'uk**	**k'aas**	**k'áax**
clay	cheek	ugly	forrest
barro	mejilla	feo	bosque
k'an	**k'a'an**	**k'iin**	**k'áan**
ripe	strong	day, sun	hammock
maduro	fuerte	día, sol	hamaca
ch'och'	**k'i'ik'**	**t'uut'**	**k'áak'**
cicada	blood	parrot	fire
cigarra	sangre	loro	fuego

Appendix B

Grammar with Fixed Hierarchical Rankings of Articulatory Constraints

B.1: Mean Ranking Values of Cue Constraints
(simultaneous production and perception learning with fixed hierarchical rankings of articulatory constraints)

	constraint	ranking value	constraint	ranking value
glottalization	*/gl/,[mod]	97.93	*/hi/,[mod]	96.86
	*/gl/,[wg]	99.09	*/hi/,[wg]	103.29
	*/gl/,[cr]	97.87	*/hi/,[cr]	102.72
	*/gl/,[gs]	93.53	*/hi/,[gs]	108.66
initial pitch	*/gl/,[in = -2]	100.38	*/hi/,[in = -2]	101.22
	*/gl/,[in = 0]	100.00	*/hi/,[in = 0]	99.29
	*/gl/,[in = 2]	98.61	*/hi/,[in = 2]	98.46
	*/gl/,[in = 4]	98.97	*/hi/,[in = 4]	99.17
	*/gl/,[in = 6]	100.11	*/hi/,[in = 6]	100.76
	*/gl/,[in = 8]	101.03	*/hi/,[in = 8]	101.94
pitch span	*/gl/,[sp = 2]	99.69	*/hi/,[sp = 2]	99.20
	*/gl/,[sp = 4]	98.91	*/hi/,[sp = 4]	98.73
	*/gl/,[sp = 6]	99.58	*/hi/,[sp = 6]	99.67
	*/gl/,[sp = 8]	100.11	*/hi/,[sp = 8]	100.71
	*/gl/,[sp = 10]	101.63	*/hi/,[sp = 10]	102.66
	*/gl/,[sp = 12]	98.59	*/hi/,[sp = 12]	100.47
articulatory	*[adducted vocal folds]		*[wg]	
			*[cr]	
			*[gs]	
	*[creaky voice, small pitch span]		*[cr, sp<6]	
			*[cr, sp<4]	
			*[cr, sp<2]	
	*[large pitch span]		*[sp>4]	
			*[sp>6]	
			*[sp>8]	
			*[sp>10]	
			*[sp>12]	

B.2: Comparison of Observed and Expected Frequencies of Input and Output Pairings
(simultaneous production and perception learning with fixed hierarchical rankings of articulatory constraints)
$\chi^2 = 612.04$
a. surface forms = /gl/

		mod			wg			cr			gs		
init pitch	pitch span	obs. freq.	exp. freq.	χ^2	obs. freq.	exp. freq.	χ^2	obs. freq.	exp. freq.	χ^2	obs. freq.	exp. freq.	χ^2
-2	2	4	1.03	1.33	0	0.07	0.48	2	0.19	0.38	0	0.03	0.23
	4	4	1.87	6.08	0	0.26	1.81	2	0.45	0.38	0	0.06	0.43
	6	1	1.04	5.25	0	0.28	1.90	1	0.48	1.57	0	0.03	0.19
	8	2	0.55	0.85	1	0.36	0.86	2	0.64	1.28	1	0.01	8.52
	10	0	0.14	0.98	1	0.11	0.10	0	0.19	1.27	0	0.00	0.02
	12	5	0.51	0.64	0	0.32	2.21	3	0.57	0.22	0	0.02	0.13
0	2	16	1.54	2.81	1	0.12	0.03	1	0.22	0.16	0	0.05	0.36
	4	5	2.47	8.38	1	0.37	0.91	2	0.59	1.02	0	0.09	0.58
	6	0	1.48	10.17	1	0.42	1.24	1	0.74	3.29	1	0.04	1.84
	8	3	0.78	1.01	2	0.49	0.54	3	0.86	1.44	0	0.02	0.11
	10	0	0.21	1.46	0	0.13	0.88	1	0.22	0.16	0	0.00	0.01
	12	0	0.62	4.24	1	0.50	1.72	3	0.76	0.95	0	0.02	0.10
2	2	20	4.07	2.21	1	0.37	0.94	3	0.60	0.30	1	0.13	0.01
	4	34	6.77	3.30	4	0.96	1.03	2	1.59	7.27	0	0.22	1.48
	6	8	3.86	12.85	7	1.18	0.14	2	1.94	9.57	0	0.12	0.83
	8	6	2.09	4.85	1	1.43	7.92	9	2.37	3.22	0	0.06	0.38
	10	1	0.54	1.94	0	0.36	2.45	2	0.60	1.10	0	0.01	0.09
	12	1	1.74	10.00	1	1.15	6.00	13	2.06	0.08	2	0.03	13.92
4	2	2	3.29	18.69	0	0.30	2.04	1	0.44	1.33	0	0.10	0.66
	4	24	5.39	4.51	4	0.78	0.34	3	1.31	3.99	0	0.15	1.02
	6	14	3.08	2.39	5	0.85	0.12	6	1.49	1.72	0	0.09	0.64
	8	5	1.69	3.75	5	1.07	0.74	3	1.78	6.91	1	0.04	2.30
	10	0	0.46	3.12	1	0.31	0.61	1	0.51	1.80	0	0.01	0.06
	12	0	1.40	9.58	2	0.95	3.11	11	1.59	0.00	0	0.04	0.25
6	2	1	1.40	7.72	0	0.10	0.68	0	0.20	1.34	0	0.05	0.34
	4	5	2.33	7.55	1	0.33	0.70	0	0.57	3.93	0	0.06	0.42
	6	10	1.30	0.13	3	0.38	0.06	4	0.64	0.03	0	0.04	0.27
	8	5	0.72	0.00	1	0.46	1.44	1	0.79	3.60	0	0.01	0.05
	10	2	0.16	0.75	1	0.13	0.02	1	0.20	0.09	0	0.00	0.03
	12	1	0.55	2.04	1	0.42	1.20	5	0.74	0.00	0	0.02	0.12
8	2	1	0.67	2.78	0	0.06	0.40	0	0.09	0.63	0	0.02	0.15
	4	2	1.03	3.62	1	0.16	0.00	0	0.24	1.62	0	0.03	0.21
	6	0	0.60	4.08	3	0.17	3.09	0	0.28	1.93	0	0.02	0.10
	8	5	0.29	4.72	2	0.20	0.29	0	0.35	2.37	0	0.00	0.03
	10	3	0.07	14.66	0	0.04	0.30	1	0.11	0.08	0	0.00	0.01
	12	0	0.24	1.66	0	0.18	1.23	0	0.30	2.05	0	0.01	0.07

b. surface form = /hi/

		mod			wg			cr			gs		
init pitch	pitch span	obs. freq.	exp. freq.	χ^2	obs. freq.	exp. freq.	χ^2	obs. freq.	exp. freq.	χ^2	obs. freq.	exp. freq.	χ^2
	2	2	1.27	5.13	0	0.01	0.05	1	0.01	14.28	0	0.00	0.00
	4	8	1.70	1.14	0	0.01	0.08	0	0.02	0.11	0	0.00	0.00
	6	2	0.93	2.99	0	0.01	0.05	0	0.02	0.13	0	0.00	0.00
	8	1	0.38	1.01	0	0.01	0.05	0	0.01	0.08	0	0.00	0.00
	10	1	0.06	0.81	0	0.00	0.01	0	0.01	0.05	0	0.00	0.00
-2	12	2	0.20	0.26	0	0.00	0.03	0	0.01	0.06	0	0.00	0.00
	2	30	6.11	3.34	0	0.02	0.14	0	0.04	0.29	0	0.00	0.00
	4	30	8.31	12.73	0	0.06	0.38	0	0.08	0.55	0	0.00	0.00
	6	5	4.42	21.10	0	0.05	0.37	0	0.10	0.66	0	0.00	0.00
	8	6	1.95	4.05	0	0.04	0.29	0	0.06	0.42	0	0.00	0.00
	10	0	0.28	1.91	0	0.00	0.01	0	0.02	0.12	0	0.00	0.00
0	12	3	0.92	1.71	0	0.02	0.14	0	0.04	0.29	0	0.00	0.00
	2	51	10.35	5.59	1	0.04	2.01	0	0.07	0.50	0	0.00	0.00
	4	56	14.49	18.87	0	0.11	0.72	0	0.15	1.04	0	0.00	0.00
	6	17	7.56	23.35	0	0.09	0.60	0	0.15	1.00	0	0.00	0.00
	8	6	3.43	13.01	0	0.08	0.53	0	0.11	0.77	0	0.00	0.00
	10	0	0.50	3.40	0	0.01	0.08	0	0.02	0.14	0	0.00	0.00
2	12	3	1.65	6.07	1	0.05	1.21	1	0.05	1.07	0	0.00	0.00
	2	13	6.59	22.89	0	0.03	0.20	0	0.05	0.33	0	0.00	0.00
	4	28	8.99	18.33	0	0.05	0.34	0	0.11	0.77	0	0.00	0.00
	6	33	4.83	0.00	0	0.06	0.40	0	0.10	0.69	0	0.00	0.00
	8	9	2.16	2.26	0	0.05	0.31	0	0.08	0.52	0	0.00	0.00
	10	1	0.32	0.64	1	0.01	8.52	0	0.01	0.05	0	0.00	0.00
4	12	0	1.11	7.60	1	0.02	4.25	2	0.03	17.05	0	0.00	0.00
	2	0	1.87	12.79	0	0.01	0.05	0	0.01	0.10	0	0.00	0.00
	4	6	2.50	7.23	0	0.02	0.11	0	0.04	0.27	0	0.00	0.00
	6	10	1.43	0.01	0	0.01	0.08	0	0.03	0.20	0	0.00	0.00
	8	6	0.63	0.68	0	0.01	0.10	0	0.03	0.19	0	0.00	0.00
	10	2	0.09	3.10	0	0.00	0.01	0	0.00	0.03	0	0.00	0.00
6	12	0	0.32	2.21	0	0.01	0.04	1	0.01	9.32	0	0.00	0.00
	2	0	0.64	4.38	0	0.00	0.01	0	0.00	0.01	0	0.00	0.00
	4	0	0.82	5.64	0	0.01	0.05	0	0.01	0.09	0	0.00	0.00
	6	3	0.48	0.02	0	0.01	0.03	0	0.01	0.06	0	0.00	0.00
	8	3	0.19	2.19	0	0.00	0.01	0	0.01	0.04	0	0.00	0.00
	10	0	0.02	0.16	0	0.00	0.00	0	0.00	0.01	0	0.00	0.00
8	12	2	0.11	2.19	0	0.00	0.03	0	0.00	0.00	0	0.00	0.00

REFERENCES

AnderBois, Scott. In press. Strong Positions and Laryngeal Features in Yukatek Maya. In *Proceedings of NELS 39*. [ROA 1021]

Anderson, Torben. 1987. The Phonemic System of Agar Dinka. *Journal of African Languages and Linguistics* 9: 1-27.

Amerman, James, Raymond Daniloff, and Kenneth Moll. 1970. Lip and Jaw Coarticulation for the Phoneme /æ/. *Journal of Speech and Hearing Research* 13: 147-161.

Avelino, Heriberto. 2004. Topics in Yalálag Zapotec, with Particular Reference to Its Phonetic Structures. Doctoral dissertation, UCLA.

Avelino, Heriberto. 2007. A Cross-Linguistic Study of Laryngeal Contrast: Phonation and Tone. Paper presented at the International Congress of Phonetic Sciences in Saarland, Germany.

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

Ayres, Glenn and Barbara Pfeiler. 1997. *Los verbos mayas: la conjugación en el maya yucateco moderno*. Mérida, Mexico: Universidad Autónoma de Yucatán.

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

Blair, Robert W. 1964. Yucatec Maya Noun and Verb Morpho-Syntax. Ph.D. dissertation. Indiana University.

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

Blevins, Juliette. 2004. *Evolutionary Phonology: The Emergence of Sound Patterns*. Cambridge: Cambridge University Press.

Blevins, Juliette and Andrew Garrett. 1998. The Origin of Consonant-Vowel Metathesis. *Language* 74: 508-56.

Blutner, Reinhard. 2000. Some Aspects of Optimality in Natural Language Interpretation. *Journal of Semantics* 17: 189-216. [ROA 389]

Boersma, Paul. 1993. Accurate Short-Term Analysis of the Fundamental Frequency and the Harmonics-To-Noise Ratio of a Sampled Sound. *Proceedings of the Institute of Phonetic Sciences*, University of Amsterdam 17: 97-110.

Boersma, Paul. 1997. How We Learn Variation, Optionality, and Probability. *Proceedings of the Institute of Phonetic Sciences*, University of Amsterdam 21: 43–58. [ROA 221]

Boersma, Paul. 1998. Functional Phonology: Formalizing the Interactions between Articulatory and Perceptual Drives. Doctoral dissertation, University of Amsterdam.

Boersma, Paul. 1999a. Optimality-Theoretic Learning in the PRAAT Program. *Proceedings of the Institute of Phonetic Sciences*, University of Amsterdam 23: 17-35. [ROA 380]

Boersma, Paul 1999b. Nasal Harmony in Functional Phonology. Ms. University of Amsterdam.

Boersma, Paul. 2006. Prototypicality Judgements as Inverted Perception. In Gisbert Fanselow, Caroline Féry, Matthias Schlesewsky & Ralf Vogel (eds.), *Gradience in Grammar*, 167-184. Oxford: Oxford University Press. [ROA 505]

Boersma, Paul. 2007a. Some Listener-Oriented Accounts of *h*-Aspiré in French. *Lingua* 117: 1989-2054. [ROA 730]

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

Boersma, Paul, Paola Escudero, and Rachel Hayes. 2003. Learning Abstract Phonological from Auditory Phonetic Categories: An Integrated Model for the Acquisition of Language-Specific Sound Categories. *Proceedings of the 15th International Congress of Phonetic Sciences*, 1013-1016. [ROA 585]

Boersma, Paul and Silke Hamann. 2008. The Evolution of Auditory Dispersion in Bidirectional Constraint Grammars. *Phonology* 25: 217-270. [ROA 909]

Boersma, Paul and Bruce Hayes. 2001. Empirical Tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32: 45-86. [ROA 348]

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

Briceño Chel, Fidencio. 2002. Lengua e identidad entre los mayas de la península de Yucatán. *Los investigadores de la cultura maya* 10: 370-379. Campeche, Mexico: Tomo II Universidad Autónoma de Campeche-SECUD. [Available online at http://www.mayas.uady.mx/articulos/lengua.html]

Briceño Chel, Fidencio. 2006. *Los verbos del maya yucateco actual*. Mérida, Mexico: Instituto Nacional de Lenguas Indígenas.

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.

Byrd, Dani. 1996. A Phase Window Framework for Articulatory Timing. *Phonology* 13: 139-69.

Campos-Astorkiza, Rebeka. 2007. Minimal Contrast and the Phonology-Phonetics Interaction. Doctoral dissertation, University of Southern California. [ROA 925]

Chen, Matthew. 1970. Vowel Length Variation as a Function of the Consonant Environment. *Phonetica* 22: 129-159.

Cho, Taehong and Peter Ladefoged. 1999. Variation and Universals in VOT: Evidence from 18 Languages. *Journal of Phonetics* 27: 207-229.

Choi, John D. 1991. An Acoustic Study of Kabardian Vowels. *Journal of the International Phonetic Association* 21: 4–12.

Chomsky, Noam and Morris Halle. 1968. *The Sound Pattern of English*. New York: Harper & Row.

Cohn, Abigail. 1990. Phonetic and Phonological Rules of Nasalization. Doctoral dissertation, UCLA.

Cohn, Abigail. 1993. Nasalization in English: Phonology or Phonetics? *Phonology* 10: 43-81.

Cohn, Abigail. 1998. The Phonetics-Phonology Interface Revisited: Where's Phonetics? *Texas Linguistics Forum* 41 (*Proceedings of the 1998 Texas Linguistics Society Conference)*: 25-40.

Cohn, Abigail. 2006. Is there Gradient Phonology? In Gisbert Fanselow, Caroline Féry, Matthias Schlesewsky & Ralf Vogel (eds.), *Gradience in Grammar*, 25-44. Oxford: Oxford University Press.

Cutler, Anne, Jacques Mehler, Dennis Norris, and Juan Segui. 1987. Phoneme Identification and the Lexicon. *Cognitive Psychology* 19: 141-177.

de Lacy, Paul. 2002. The Formal Expression of Markedness. Doctoral dissertation, University of Massachusetts, Amherst. [ROA 542]

Diccionario Maya Popular, 2nd ed. 2004. Mérida, Mexico: Academia de la Lengua Maya de Yucatán.

Dresher, B. Elan. In press. *The Contrastive Hierarchy in Phonology.* Cambridge: Cambridge University Press.

Edmondson, Jerold A. and John H. Esling. 2006. The Valves of the Throat and their Functioning in Tone, Vocal Register, and Stress: Laryngoscopic Case Studies. *Phonology* 23: 157-191.

Escudero, Paola. 2001. The Role of the Input in the Development of L1 and L2 Sound Contrasts: Language-Specific Cue Weighting for Vowels. In Anna H.-J. Do, Laura Dominguez and Aimee Johansen (eds.), *Proceedings of the 25th Annual Boston University Conference on Language Development,* 250-261. Somerville, MA: Cascadilla.

Escudero, Paola. 2002. The Perception of English Vowel Contrasts: Acoustic Cue Reliance in the Development of New Contrasts. In Jonathan Leather and Allan James (eds.), *Proceedings of the 4th Symposium on the Acquisition of Second Language Speech, New Sounds 2000,* 122-131. University of Klagenfurt.

Escudero, Paola and Paul Boersma. 2004. Bridging the Gap between L2 Speech Perception Research and Phonological Theory. *Studies in Second Language Acquisition*, 26: 551-585.

Fischer-Jørgensen, Eli. 1989. Phonetic Analysis of the Stød in Standard Danish. *Phonetica* 46:1-59.

Fisher, William M. 1973. Towards the Reconstruction of Proto-Yucatec. Doctoral dissertation, University of Chicago.

Fisher, William M. 1976. On Tonal Features in the Yucatecan Dialects. In *Mayan Linguistics: Volume One*, ed. Marlys McClaran, 29-43. Los Angeles: American Indian Studies Center.

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

Foulkes, Paul. 2006. Phonological Variation: A Global Perspective. In Bas Aarts and April M. S. McMahon (eds.), *The Handbook of English Linguistics*, 625-669. Malden, MA: Blackwell.

Fourakis, Marios and Robert Port. 1986. Stop Epenthesis in English. *Journal of Phonetics* 14: 197-221.

Frazier, Melissa. 2006. Output-Output Faithfulness to Moraic Structure: Evidence from American English. In Christopher Davis, Amy Rose Deal, and Youri Zabbal (eds.), *Proceedings of NELS* 36, v.1, 1-14. [ROA 820]

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

Frazier, Melissa. In press. A Bidirectional Stochastic OT Analysis of Production and Perception in Yucatec Maya. In *Proceedings of NELS 39*.

Gordon, Matthew and Peter Ladefoged. 2001. Phonation Types: A Cross-Linguistic Overview. *Journal of Phonetics* 29: 383-406.

Gordon, Raymond G., Jr. (ed.). 2005. Ethnologue: Languages of the World, 15th edition. Dallas: SIL International. Online version: http://www.ethnologue.com/.

Gouskova, Maria. 2003. Deriving Economy: Syncope in Optimality Theory. Doctoral dissertation, University of Massachusetts, Amherst. [ROA 610]

Güémez Pineda, Miguel. 1994. La lengua maya en Yucatán: una perspectiva sociodemográfica. *I'inaj,* Semilla de Maíz, Revista de divulgación del patrimonio cultural de Yucatán, Conaculta, Instituto Nacional de Antropología e Historia. [Available online at http://www.mayas.uady.mx/articulos/miguel.html]

Gussenhoven, Carlos and Renske Teeuw. 2008. A Moraic and a Syllabic H-tone in Yucatec Maya. In E. H. Z & P. M. Butrageno (eds.), *Fonología instrumental: patrones fónicos y variacion*, 49-71. Mexico City: Colegio de México.

Hanks, William F. 1984. Sketch Grammar of Yucatec Maya. Ms. Department of Anthropology, University of Chicago.

Hombert, Jean-Marie. 1978. In Victoria A. Fromkin (ed.), *Tone: A Linguistic Survey*, 77-111. New York: Academic Press.

Hoogshagen, Searle. 1959. Three Contrastive Vowel Lengths in Mixe. *Zeitschrift für Phonetik und Allgemeine Sprachwissenschaft* 12: 111-115.

Hyman, Larry. 1977. Phonologization. In Alphonse G. Juilland (ed.), *Linguistic Studies Offered to Joseph Greenberg on the Occasion of his Sixtieth Birthday*, 407-18. Saratoga, CA: Alma Libri.

Jassem, Wiktor. 2003. Polish. *Journal of the International Phonetic Association* 33: 103–107.

Kaufman, Terrence. 1976. Archaeological and Linguistic Correlations in Mayaland and Associated Areas of Meso-America. *World Archaeology* 8: 101-118.

Kaufman, Terrence. 1984. Cross-Currents of Grammatical Innovation in the Mayan Languages, with Special Reference to Tojolabal: Intro-Mayan Diffusion. Ms. Department of Anthropology, University of Pittsburg.

Keating, Patricia A. 1985a. CV Phonology, Experimental Phonetics, and Coarticulation. *UCLA Working Papers in Phonetics* 62: 1-13.

Keating, Patricia A. 1985b. Universal Phonetics and the Organization of Grammars. In Victoria A. Fromkin (ed.), *Phonetic Linguistics*, 115-132. New York: Academic Press.

Keating, Patricia A. 1988. The Phonology-Phonetics Interface. In Frederick J. Newmeyer (ed.), *Linguistics: The Cambridge Survey, v. 1: Grammatical Theory*, 281-302. Cambridge: Cambridge University Press.

Keating, Patricia A. 1990. The Window Model of Coarticulation: Articulatory Evidence. In John Kingston and Mary Beckman (eds.), *Papers in Laboratory Phonology* I: 451-470. Cambridge: Cambridge University Press.

Kingston, John. 2007. The Phonetics-Phonology Interface. In Paul de Lacy (ed.), *The Handbook of Phonology*, 401-434. Cambridge: Cambridge University Press.

Kirchner, Robert. 1997. Contrastiveness and Faithfulness. *Phonology* 14: 83-111.

Kirchner, Robert. 1998. An Effort Based Approach to Consonant Lenition. Doctoral dissertation, UCLA. [ROA 276]

Kirk, Paul L., Jenny Ladefoged, and Peter Ladefoged. 1993. Quantifying Acoustic Properties of Modal, Breathy, and Creaky Vowels in Jalapa Mazatec. In Anthony Mattina and Timothy Montler (eds.), *American Indian Linguistics and Ethnography in Honor of Lawrence C. Thompson*, 435-450. Ann Arbor: University of Michigan.

Krämer, Martin. 2001. Yucatec Maya Vowel Alternations: Harmony as Syntagmatic Identity. *Zeitschrift für Sprachwissenschaft* 20: 175-217. [ROA 423]

Krauss, Michael. 1992. The World's Languages in Crisis. *Language* 68: 4-10.

Kügler, Frank, Stavros Skopeteas, and Elisabeth Verhoeven. 2007. Encoding Information Structure in Yucatec Maya: On the Interplay of Prosody and Syntax. *Interdisciplinary Studies on Information Structure* 8: 187-208.

Ladd, D. Robert. 1996. *Intonational Phonology*. Cambridge: Cambridge University Press.

Ladefoged, Peter. 1971. *Preliminaries to Linguistic Phonetics*. Chicago: University of Chicago Press.

Ladefoged, Peter. 1972. Phonetic Prerequisites for a Distinctive Feature Theory. In Albert Valdman (ed.), *Papers in Linguistics and Phonetics to the Memory of Pierre Delattre*, 273-286. The Hague: Mouton.

Laver, John. 1980. *The Phonetic Description of Voice Quality*. Cambridge: Cambridge University Press.

Lehmann, Christian. 1998. *Possession in Yucatec Maya*. Munich: Lincom Europa.

Liberman, Alvin M., Katherine Safford Harris, Howard S. Hoffman, and Belver C. Griffith. 1957. The Discrimination of Speech Sounds within and across Phoneme Boundaries. *Journal of Experimental Psychology* 54: 358-368.

Liberman, A. M., P. C. Delattre, and F. S. Cooper. 1958. Some Cues for the Distinction between Voiced and Voiceless Stops in Initial Position. *Language and Speech* 1: 153-167.

Lisker, Leigh and Arthur S. Abramson. 1964. Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements. *Word* 20: 384-422.

Lisker, Leigh and Arthur S. Abramson. 1970. The Voicing Dimension: Some Experiments in Comparative Phonetics. In Bohuslav Hála, Milan Romportl, and Přemysl Janota (eds.), *Proceedings of the Sixth International Congress of Phonetic Sciences*, 563-567.

Magri, Giorgio. 2008. Linear Methods in Optimality Theory: a Convergent Incremental Algorithm that Performs both Promotion and Demotion. Paper presented at NecPhon2, Yale, Nov. 15. (downloaded from http://web.mit.edu/gmagri/www/papers/LinearMethodsOT[2].pdf on 03/31/09)

McCarthy, John J. 2002. *A Thematic Guide to Optimality Theory*. Cambridge: Cambridge University Press.

McCarthy, John J. and Alan Prince. 1993. Generalized Alignment. *Yearbook of Morphology*: 79–153. [ROA 7]

McCarthy, John J. and Alan Prince. 1995. Faithfulness and Reduplicative Identity. In Jill Beckman, Suzanne Urbanczyk, and Laura Walsh Dickey (eds), *University of Massachusetts Occasional Papers in Linguistics* 18: *Papers in Optimality Theory*, 249-384. Graduate Linguistic Student Association, University of Massachusetts, Amherst. [ROA 60]

McQuown, Norman A. 1975. El acento del maya yucateco. *Lingüística e indigenismo moderno de América* 5: 59-71.

Miller, Wick R. 1965. *Acoma Grammar and Texts*. Berkeley: University of California Press.

Myers, Scott. 2002. Gaps in Factorial Typology: The Case of Voicing in Consonant Clusters. Ms. University of Texas, Austin. [ROA 509]

Ní Chasaide, Ailbhe and Christer Gobl. 1995. Voice Source Variation. In William J. Hardcastle and John Laver (eds.), *The Handbook of Phonetic Sciences*, 427-61. Cambridge, MA: Blackwell.

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

Ohala, John J. 1981. The Listener as a Source of Sound Change. In Carrie S. Masek, Robert A. Hendrick and Mary Frances Miller (eds.), *Papers from the Parasession on Language and Behavior*, Chicago Linguistic Society, 178-203.

Ohala, John J. 1993. The Phonetics of Sound Change. In Charles Jones (ed.), *Historical Linguistics: Problems and Perspectives*, 237-278. London: Longman.

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

Pater, Joe. 2004. Bridging the Gap between Receptive and Productive Development with Minimally Violable Constraints. In René Kager, Joe Pater, and Wim Zonneveld (eds.), *Constraints in Phonological Acquisition*, 219-244. Cambridge: Cambridge University Press.

Pater, Joe. 2008. Gradual Learning and Convergence. *Linguistic Inquiry* 39: 334-345. [ROA 917]

Peterson, Gordon E. and Harold L. Barney. 1952. Control Methods Used in a Study of the Vowels. *Journal of the Acoustical Society of America* 24: 175–184.

Peterson, Gordon E. and Ilse Lehiste. 1960. Duration of Syllable Nuclei in English. *Journal of the Acoustical Society of America* 32: 693–703.

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

Pike, Kenneth L. 1946. Phonemic Pitch in Maya. *International Journal of American Linguistics* 12: 82-88.

Plauché, Madelaine C., Rosemary Beam de Azcona, Rungpat Roengpitya, and William F. Weigel. 1998. Glottalized Sonorants: A Phonetic Universal? In B. K. Bergen, M. C. Plauché, and A. C. Bailey (eds.), *Proceedings of the 24th Annual Meeting of the Berkeley Linguistics Society*, 381-390. Berkeley Linguistics Society.

Prince, Alan S. 1980. A Metrical Theory for Estonian Quantity. *Linguistic Inquiry* 11: 511-562.

Prince, Alan S. 1997a. Paninian Relations. Colloquium talk at University of Massachusetts, Amherst. (downloaded from http://equinox.rutgers.edu/gamma/talks/umass1997.pdf on 03/27/09)

Prince, Alan S. 1997b. Stringency and Anti-Paninian Hierarchies. Handout from LSA Institute. (downloaded from http://equinox.rutgers.edu/gamma/talks/insthdt2.pdf on 03/27/09)

Prince, Alan S. and Paul Smolensky. 1993/2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Malden, MA: Blackwell. [Revision of 1993 technical report, Rutgers University Center for Cognitive Science; ROA 537]

Redi, Laura and Stefanie Shattuck-Hufnagel. 2001. Variation in the Realization of Glottalization in Normal Speakers. *Journal of Phonetics* 29: 407-429.

Remijsen, Bert and Leoma Gilley. 2008. Why are Three-Level Vowel Length Systems Rare? Evidence from Dinka (Luanyjang Dialect). *Journal of Phonetics* 36: 318-344.

Roffe, G. Edward. 1946. The Phonemic Structure of Lao. *Journal of the American Oriental Society* 66: 289-295.

Silverman, Daniel. 1997. Laryngeal Complexity in Otomanguean Vowels. *Phonology* 14: 135-162.

Silverman, Daniel, Barbara Blankenship, Paul Kirk and Peter Ladefoged. 1995. Phonetic Structures in Jalapa Mazatec. *Anthropological Linguistics* 37: 70-88.

Smolensky, Paul. 1996. On the Comprehension/Production Dilemma in Child Language. *Linguistic Inquiry* 27: 720–731. [ROA 118]

Terán Contreras, Silvia and Christian H. Rasmussen. 2005. *Xocén: El pueblo en el centro del mundo*. Mérida, Mexico: Universidad Autónoma de Yucatán.

Tesar, Bruce. 1997. An Iterative Strategy for Learning Metrical Stress in Optimality Theory. In Elizabeth Hughes, Mary Hughes & Annabel Greenhill (eds.), *Proceedings of the 21st Annual Boston University Conference on Language Development*, 615–626. Somerville, Mass.: Cascadilla. [ROA 117]

Tesar, Bruce & Paul Smolensky. 1998. Learnability in Optimality Theory. *Linguistic Inquiry* 29: 229–268. [ROA 155]

Tesar, Bruce & Paul Smolensky. 2000. *Learnability in Optimality Theory*. Cambridge, Mass.: MIT Press.

Tsunoda, Tasaku. 2005. *Language Endangerment and Language Revitalization*. New York: Mouten de Gruyter.

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

Ward, Paul H., Jay W. Sanders, Ronald Goldman, and G. Paul Moore. 1969. Diplophonia. *Annals of Otology, Rhinology, and Laryngology* 78: 771-777.

Wichmann, Søren. 1995. *The Relationship among the Mixe-Zoquean Languages of Mexico*. Salt Lake City: University of Utah Press.

Wilson, Colin. 2001. Consonant Cluster Neutralisation and Targeted Constraints. *Phonology* 18: 147-197.

Witschey, Walter R. T. 2009. Fig. 1.1, Map of Yucatan, extracted from the Electronic Atlas of Ancient Maya Sites. http://MayaGIS.smv.org. Personal communication.

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

Zeevat, Henk. 2000. The Asymmetry of Optimality Theoretic Syntax and Semantics. *Journal of Semantics* 17: 243-262.

Zsiga, Elizabeth C. 1997. Feature, Gestures, and Igbo Vowel Assimilation: An Approach to the Phonology/Phonetics Mapping. *Language* 73: 227-274.