**Title**
Production and perception of glottal stops

**Permalink**
https://escholarship.org/uc/item/7zk830cm

**Author**
Garellek, Marc

**Publication Date**
2013

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

# Production and perception of glottal stops

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Linguistics

by

## Marc Garellek

2013

ABSTRACT OF THE DISSERTATION

# Production and perception of glottal stops

by

## Marc Garellek

Doctor of Philosophy in Linguistics

University of California, Los Angeles, 2013

Professor Patricia A. Keating, Chair

This dissertation investigates how glottal stops are produced and perceived, and why they occur so frequently before word-initial vowels in languages of the world. Specifically, the goal of the production chapter was to determine whether glottal stops are truly glottal sounds. High-speed imaging of glottal stops uttered by five phonetically-trained English speakers was obtained using trans-oral videoendoscopy. When produced as plosives, glottal stops always had some form of vocal fold incursion, though full glottal closure is not always observed. Moreover, glottal stops were usually – but not always – accompanied by ventricular incursion. Glottal stops are thus necessarily glottal but not necessarily ventricular sounds. The timing of ventricular incursion suggests that it may be used to sustain, rather than produce, glottal closure.

The specific goal of the perception study was to determine whether glottal stops are perceived distinctly from phrase-final creak, which shares similar acoustic features to glottalization. Sixteen English listeners were asked to identify words with glottal stops as allophones of /t/, e.g. *button, atlas*, and *dent*. They were also presented with near-minimal pairs with no glottal stop, e.g. *bun, Alice*, and *den*. The target words either had creak or no creak. The results indicate that words with glottal stop allophones are more accurately and more confidently identified when no creak is present, when the glottal stop is acoustically longer, and/or when the glottal stop is word-medial. Moreover, glottal stops are more accurately identified in creak when they occur in *button*-type words, compared with worse identification

in *atlas-* and *dent*-type words. For words with no glottal stops (e.g., *bun, Alice,* and *den*), the presence of creak does not render them confusable with words with glottal allophones, except for *den*-type words where word-final creak is sometimes mistaken for a glottal stop. Thus, glottal stops are generally harder to detect in creak, but are mutually confusable with creak only word-finally after nasals.

To determine why and when glottal stops occur word-initially, the occurrence of word-initial full glottal stops in an English corpus was analyzed using logistic mixed-effects regression modeling. Prominence and phrasing are overwhelmingly the most important factors in predicting full glottal stop occurrence in English. Moreover, prominent word-initial vowels that are not preceded by full glottal stops show acoustic correlates of glottal constriction, whereas non-prominent phrase-initial vowels do not. Rather, phrase-initial voicing (even for sonorants) is less regular, but in a manner inconsistent with glottal constriction. These findings were subsequently confirmed using articulatory measures from electroglottography, and extended to Spanish. Based on the results, a prominence-driven theory of word-initial glottalization is proposed and motivated, with higher phrasal domains responsible only for the relative strength of the glottal stop gesture. Glottalization before word-initial vowels is thus a marker of prominence and is used amplify cues to prominence when they would otherwise be weakened.

The dissertation of Marc Garellek is approved.

Abeer Alwan

Sun-Ah Jun

Jody Kreiman

Megha Sundara

Patricia A. Keating, Committee Chair

University of California, Los Angeles

2013

TABLE OF CONTENTS

# List of Figures

# LIST OF TABLES

Being part of various voice projects at UCLA has been extremely useful in helping me shape ideas that appear in this dissertation. Thank you Pat, Jody, and Abeer for including me, and to Gang, Christina Esposito, Sameer ud Dowla Khan, Jianjing Kuang, and Kristine Yu for being great partners in creak.

Thanks to my friends and family at UCLA and back home, who forgave me when I said [aɪm ɹʌɪɾɪŋ maɪ dɪsɚˑtʰeɪʃn̩ ɒn ʔ], and to UCSD's Linguistics Department for fueling me with a deadline. Finally, I cannot thank Danny enough for reminding me to be human every now and then.

| | |
|---|---|
| 1986 | Born, Montreal, Quebec. |
| 2008 | B.A. with First-Class Honours (Linguistics) |
| | McGill University |
| | Montreal, Quebec |
| 2010 | M.A. (Linguistics) |
| | University of California, Los Angeles |
| | Los Angeles, California |
| 2010–2011 | Teaching Assistant |
| | Linguistics Department |
| | University of California, Los Angeles |
| 2010–2012 | Research Assistant |
| | Linguistics Department |
| | University of California, Los Angeles |
| 2010–2013 | Research Assistant |
| | Department of Head and Neck Surgery |
| | David Geffen School of Medecine |
| | University of California, Los Angeles |

## PUBLICATIONS

Garellek, M. (2011). Lexical effects on English vowel laryngealization. *Proceedings of the 17th International Congress of Phonetic Sciences*, 727–730.

Keating, P., Esposito, C. M., Garellek, M., Khan, S., & Kuang, J. (2011). Phonation contrasts across languages. *Proceedings of the 17th International Congress of Phonetic Sciences*, 1046–1049.

Daland, R., Hayes, B., White, J., Garellek, M., Davis, A., & Normann, I. (2011). Explaining sonority projection effects. *Phonology, 28*, 197–234.

Garellek, M. & Keating, P. (2011). The acoustic consequences of phonation and tone interactions in Jalapa Mazatec. *Journal of the International Phonetic Association, 41*, 185–205.

Garellek, M. (2012). The timing and sequencing of coarticulated non-modal phonation in English and White Hmong. *Journal of Phonetics, 40*, 152–161.

Zhang, Z., Kreiman, J., Gerratt, B. R., & Garellek, M. (2013). Acoustic and perceptual effects of changes in body-layer stiffness in symmetric and asymmetric vocal fold models. *Journal of the Acoustical Society of America, 133*, 453–462.

Garellek, M., Keating, P., Esposito, C. M., & Kreiman, J. (2013). Voice quality and tone identification in White Hmong. *Journal of the Acoustical Society of America, 133*, 1078–1089.

# CHAPTER 1

# Introduction

## 1.1 Overview of the phonetics of glottal stops

In this dissertation, I investigate how glottal stops are produced and perceived, and when and why they occur word-initially. Glottal stops occur frequently across languages. They are attested in the phoneme inventories of nearly half the world's languages (Maddieson, 1984), yet their phonetic characteristics remain understudied. Moreover, there are several unique phonetic and distributional peculiarities about glottal stops: remarkably, it is still unclear whether glottal stops are really glottal, really stops, and why they occur 'optionally' before word-initial vowels.

### 1.1.1 Are glottal stops really *glottal*?

The first peculiarity regarding glottal stops relates to their definition as a solely glottal segment. Visual observation of the larynx during the production of glottal stops has revealed that they often – if not always – involve movement of supraglottal structures: the ventricular folds, the aryepiglottic folds and epiglottis, and even pharyngeal constriction (Esling, 1996, 1999; Harris, 1999; Zawaydeh, 1999; Harris, 2001; Esling, Fraser & Harris, 2005; Esling, Zeroual & Crevier-Buchman, 2007; Hassan & Esling, 2007; Moisik & Esling, 2011; Edmondson, Chang, Hsieh & Huang, 2011).

If glottal stops are never truly 'glottal,' why do phoneticians call them such? Indeed, Esling et al. (2007) state (p. 585) that 'a "glottal stop" cannot be uniquely glottal but is at least glottoventricular,' meaning that the ventricular folds (structures superior to the vocal

folds in the larynx) are always involved in glottal stop articulation. But as I discuss in the following section, glottal stops are known to be very variable in their acoustic realization, ranging from full stops to laryngealized phonation. Is it then possible that *some* glottal stops are really glottal, whereas others additionally involve structures above the glottis?

### 1.1.2 Are glottal stops really *stops*?

Another interesting fact about glottal stops is that they are known to be extremely variable in their production; they usually appear as a form of laryngealized voicing (Pierrehumbert & Talkin, 1992; Dilley, Shattuck-Hufnagel & Ostendorf, 1996), and listeners seem to rely predominantly on such changes in voice quality when perceiving them (Hillenbrand & Houde, 1996). Therefore, it appears as though there is nothing 'stop-like' required for the sound, both in terms of acoustic realization and in terms of perception.

If – as previous work has shown – glottal stops are cued by irregular changes in voicing (such as we see in laryngealized or creaky voicing), then they should be confusable with other linguistic gestures that involve irregular voicing, e.g. phrase-final creak.

### 1.1.3 Why are glottal stops so common word-initially?

Another curiosity is that most languages tend to 'insert' glottal stops before vowel-initial words, at least in strong prosodic positions, e.g. phrase-initially (Pierrehumbert & Talkin, 1992; Dilley et al., 1996). Glottal stops before word-initial vowels are often optional in a language, though for some languages they are obligatory (Hayes, 2009). No other phonological insertion rule seems to be as widespread as glottal stop insertion across languages, and it is certainly puzzling that all languages which insert a consonant before vowel-initial words should insert the same sound.

Although there have been many studies of the factors that influence glottal stop occurrence, there has been little investigation of which factors are most important for promoting word-initial glottal stops. That is, we know what factors play a role in their occurrence, but

we do not know how important they are compared to each other. However, it is especially important to determine which are most important, so that we can understand why glottal stop 'insertion' occurs.

## 1.2 Research questions and overview

As outlined above, there are still many unanswered questions regarding the variability in glottal stops' production, perception, and occurrence. The overarching questions that I will address in this work are therefore the following:

1. How are glottal stops produced?

2. Do listeners perceive glottal stops as distinct from creaky voice?

3. When do glottal stops occur word-initially?

4. Why are glottal stops so common before vowel-initial words?

To investigate how glottal stops are produced, in Chapter 3 I report the results of a high-speed imaging study whose goal is to determine how often glottal stops are produced as full stops, and how often the ventricular folds are involved in their production. In Chapter 4, I study glottal stop perception. In particular, I look at the effect of creaky voice on the perception of glottal stops, to determine whether listeners do indeed only rely on cues shared by both creaky voice and glottal stops when perceiving the latter. Chapter 5 is devoted to the question 'when and why do glottal stops occur before word-initial vowels?' First I analyze an English corpus to determine what factors best predict the occurrence of glottal stops. Then, using both acoustic (English) and articulatory measures (English and Spanish), I study the role of these factors in influencing voice quality for vowels with no glottal stops. Finally, I synthesize the qualitative corpus and voice quality results to develop a theory of when and

why word-initial glottal stops occur. Chapter 6 contains a general discussion and conclusions on the production and perception of glottal stops, with ideas for subsequent research.

## 1.3  Definitions of terms used in the dissertation

Before I proceed to a survey of relevant literature on glottal stops (Chapter 2), some terminological clarifications should be made.

### 1.3.1  Glottal stop

A glottal stop generally refers to a stop made in the larynx by the abrupt and sustained closure of the vocal folds. The closure is called 'sustained' to distinguish the glottal stop gesture from the periodic closing of the vocal folds during phonation (see Section 2.2.3). In this dissertation, a glottal stop refers to an abstract articulatory target (either phonemic or allophonic). Glottal stops may be phonetically realized in different ways, as will be discussed in detail below, and so in this work a glottal stop does not represent a single articulation (i.e. a phone) but rather a set of articulations ranging from laryngealized phonation, to sustained glottal closure (with or without additional supraglottal constriction). On the other hand, when referring to actual phones, I will make use of the following terms:

- 'Full glottal stop' = a glottal stop produced with full, sustained vocal fold closure: [ʔ]

- 'Incomplete glottal stop' = a glottal stop realized with incomplete vocal fold closure (i.e., as laryngealized voice): [ʔ̰]

- 'Laryngealization' = either [ʔ̰], or the kind of voice quality which involves increased vocal fold constriction usually (though not in this work) transcribed with [ ̰] (see Section 2.1.2)

- 'Creak' = irregular voice quality that does not necessarily involve increased vocal fold constriction, such as phrase-final creak (Slifka, 2000, 2006): [ ̰]. Note that this definition

differs from that given by Laver (1980) and Ladefoged & Maddieson (1996), for whom 'creak' and laryngealization are largely synonymous, with both characterized by an increase in glottal closure (see Laver (1980, p. 126)).

Note that I make a distinction between two types of irregular voicing: laryngealization and creak. This is because they differ articulatorily (Slifka, 2006), and because incomplete glottal stops are expected to have voicing that is similar articulatorily and acoustically to laryngealization, but not necessarily to creak.

### 1.3.2   Glottalization

Another source of confusion in the literature lies with the definition of glottalization. From an articulatory point of view, glottalization refers strictly to the addition of an articulatory glottal stop (i.e., full and sustained vocal fold adduction or [ʔ]). The addition of a glottal articulation to oral stops is also known as 'glottal reinforcement' (Higginbottom, 1964; Esling et al., 2005), especially for English. However, these sounds are also referred to simply as 'glottalized.' Other sounds can also have a secondary glottal articulation, including sonorants (Esling et al., 2005; Bird, Caldecott, Campbell, Gick & Shaw, 2008) and clicks (Miller, 2007). Strictly speaking, glottalization refers only to complete vocal fold adduction that may accompany a sound as a secondary articulation, and thus makes no reference to the phonation of voiced segments adjacent to the glottal stop, which are often laryngealized as the vocal folds prepare for glottal closure. However, such laryngealized phonation is also often called glottalization, especially when referring to its acoustic output (Huffman, 2005). Certain researchers have found it useful to have a term that covers phonetic effects seen for both glottal closure and laryngealized phonation, regardless of the target gesture for which the speaker aims (Henton, Ladefoged & Maddieson, 1992; Michaud, 2004; Brunelle, Nguyễn & Nguyễn, 2010). In this dissertation, I too use the term 'glottalization' when referring to the articulatory or acoustic effects on targets of either glottal stop or laryngealization.

### 1.3.3  Glottal stop vs. glottal attack

Researchers, especially those who investigate the effects of different singing styles on the voice, have studied different methods of voicing initiation, which is often termed glottal or vocal attack (Moore, 1938; Gay, Hirose, Strome & Sawashima, 1972; Hirose & Gay, 1973; Cooke, Ludlow, Hallett & Selbie, 1997; Baken & Orlikoff, 2000; Orlikoff, Deliyski, Baken & Watson, 2009; Freeman, Woo, Saxman & Murry, 2012).  In these works, voicing initiation that is preceded by complete adduction of the vocal folds (and typically ventricular adduction) is often referred to as a hard glottal attack. Thus, a hard glottal attack can be thought of as an articulatorily equivalent to a glottal stop, though the latter term is more common in linguistics than in singing or speech science. Note also that by definition a hard glottal attack is a form of voicing initiation, whereas a glottal stop is often used as a form of voicing cessation, and is not always tied to the beginning of an utterance.[1]

In sum, the main definitions to keep in mind are the following:

1. *glottal stop* = target articulation, may be 'full' [ʔ] or 'incomplete' [ʔ̰]

2. *laryngealized phonation* (or *laryngealization*) = in this dissertation, either the voicing during [ʔ̰] or when referring to certain contrastive/allophonic voice qualities in other languages

3. *glottalization* = glottal stop or laryngealization from a nearby glottal stop

4. *glottal attack* = phonation onset (will only be used when citing sources which use the term)

---

[1]Other common forms of glottal attack include the breathy attack (articulatory equivalent to [h]), and a normal (sometimes called a soft or simultaneous) attack, which is articulatorily equivalent to a state of prephonation where the vocal folds are approximated and stiffened slightly (Harris, 2001), as is common for an unaspirated stop. Other types of attack in singing include the staccato and German gestures (Freeman et al., 2012).

# CHAPTER 2

# Background on glottal stops

This chapter provides relevant background on glottal stops' production, perception, and distribution. It is structured as follows: in Section 2.1, I discuss glottal constriction gestures that occur in languages of the world, focusing on glottal stops and laryngealized phonation. In Section 2.2, I review the relevant laryngeal anatomy and physiology. Then in Section 2.3, the phonetic literature on glottal stops and other forms of glottalization is reviewed. In Section 2.4, I summarize the previous findings and motivate the studies that comprise the remaining chapters of the dissertation.

## 2.1 Glottal gestures across languages

Glottal stops form part of the phonemic inventory of 47.9% of languages in the the UCLA Phonological Segment Inventory Database or UPSID (Maddieson, 1984; Maddieson & Precoda, 1990), and of 42.6% of the languages in the Lyon-Albuquerque Phonological Systems Database (LAPSyD; Maddieson et al. (2011)). The areal distribution of glottal stops is shown in Figure 2.1 (p. 8). The glottal stop /ʔ/ is less common across languages than the voiceless glottal fricative /h/, which is found in 64.8% of the languages in UPSID. The presence of /ʔ/ in a language usually implies the presence of /h/ as well; only 18.1% of languages with glottal stops in their segmental inventory do not have /h/.

However, in many other languages (e.g. Arabic, English, German, and Ilokano), glottal stops can be found word-initially, either as an obligatory or optional allophone. As mentioned

earlier, this phenomenon is called glottal stop insertion (Hayes, 2009), and the reasons why it occurs – and why it occurs in so many languages – remain to be understood.
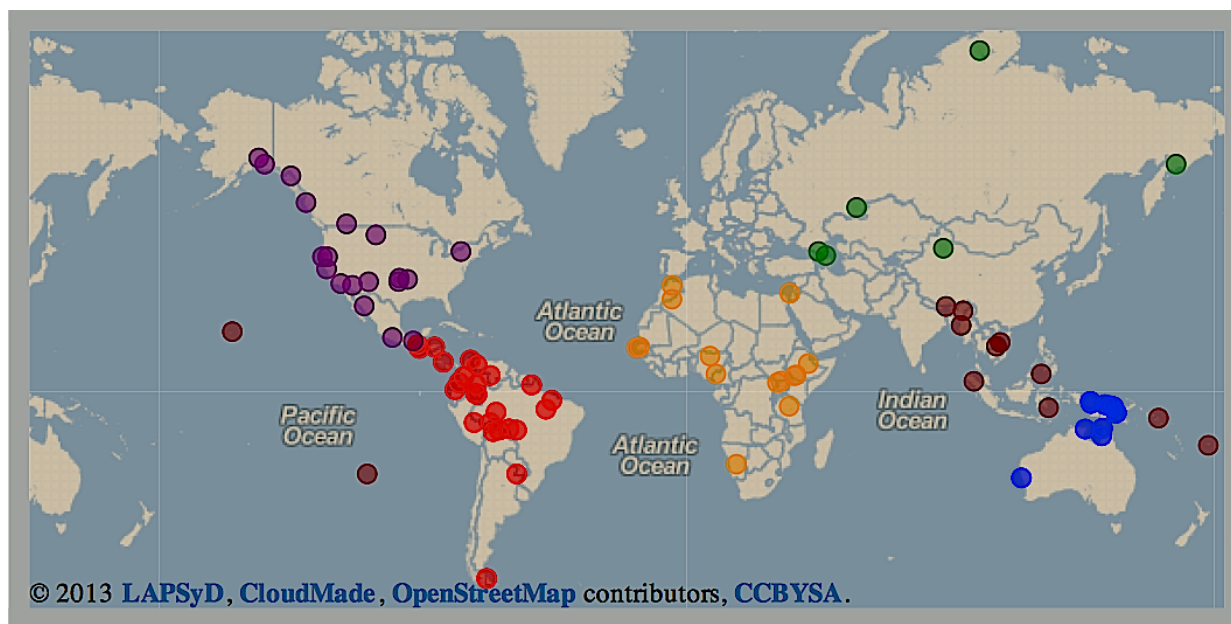


Figure 2.1: Areal distribution of languages with glottal stop phonemes in LAPSyD (Maddieson et al., 2011). The map is generated using LAPSyD's interactive website, which allows users to search for languages with /ʔ/ and generate a map with the approximate central location of speakers for each language marked by a circle. The color of the circle (online) differentiates language groups roughly by continent. The online search was generated on April 26, 2013.

UPSID also has three other types of glottal stops listed, though they are very rare. A voiced glottal plosive, which I denote by /ʔ̬/, is attested for one language, Nenets (Uralic, Russia). However, the sound is elsewhere described as a nasalized glottal stop /ʔ̃/ (Comrie, 1981, p. 117; Janhunen, 1986). Unlike other stops, [ʔ] per se can never be voiced, because adducted vocal folds cannot vibrate periodically to produce voicing. Thus, languages with voicing contrasts in stops typically only have the voiceless /ʔ/ at the laryngeal place of articulation. Gimi (Trans-New Guinea, Papua New Guinea), has a voicing contrast even for glottal stops, but the voiced glottal 'stop' is realized instead as a laryngealized glottal approximant (Ladefoged & Maddieson, 1996, p. 76).

Further, a pharyngealized glottal plosive /ʔˤ/ is attested in two languages in the database, Rutul (East Caucasian, Caucasus) and Tseshaht (also known as Nuuchahnulth). A glottal stop may also be labialized (/ʔʷ/), as is found in only one language in UPSID, Kabardian (West Caucasian, Caucasus). The presence of one of these rare glottal plosives (either of /ʔ, ʔˤ, ʔʷ/) implies the presence of a plain /ʔ/, at least for the languages included in that database.

Phonologically, glottal stops and other forms of glottalization have been grouped together using the feature [+constricted glottis] (Halle & Stevens, 1971) or [glottalization] (Lombardi, 1991), though for a more recent proposal based on visual observation, see Moisik & Esling (2011). Additionally, the glottal stop (and the epiglottal stop) need not be specified for nasality (unlike other stops), since the airstream meets the glottal closure before the velo-pharyngeal port. Thus, in the case of Nenets, although the 'nasalized' glottal stop might be so called because phonologically it patterns with nasals, during glottal closure there is no air flowing through the nose. The fact that glottal stops are unspecified for nasality is evident in some language alternations, e.g. in the Maipuran language Terena of Brazil (Bendor-Samuel, 1970; see also discussion by Laver, 1994, p. 212).

Glottal stops are also often considered placeless phonologically, because they involve little or no constriction in the vocal tract. However, in some Semitic languages they pattern with the class of pharyngeals (McCarthy, 1994). Possible phonetic motivation for this is discussed in Section 2.2.4. For further information regarding the phonological patterning of glottal stops, see Lombardi (2002) and Borroff (2007).

### 2.1.1 Glottal stop vs. zero

Perhaps because of the difficulty in producing regular phonation phrase-initially, it is quite rare for languages to have a contrast between /ʔ/ (which is likely to be realized with irregular voicing) and ∅ in onsets. Languages in which /#ʔV/ contrasts with /#V/ include some Malayo-Polynesian languages, e.g. Tongan, Tahitian, and Samoan, and Harris (2001) cites

two other such languages, both also Malayo-Polynesian: Nga'da and Mungaba Rennellese. Additionally, this contrast is said to be marginal in some Mayan languages: e.g. see Lichtman, Chang, Cramer, del Rio, Hallett, Huensch & Morales (2010) for Q'anjob'al.

Thus, many but not all languages allow for glottal stops to be inserted at the onset of vowel-initial words. The fact that glottal stop insertion before vowels is so widespread suggests that a common feature of language or speech is responsible for its occurrence. Phonologically, glottal stop insertion can be motivated by the cross-linguistic tendency for a syllable to begin with an onset (see, e.g., Lombardi (2002)), though such accounts normally do not model the optionality of glottal stop insertion.

### 2.1.2 Glottal stop vs. laryngealization

Phonemic laryngealized phonation is often thought to be derived historically from glottal stop lenition, e.g. in the Popolocan language Mazatec, spoken in Mexico (Silverman, 1995). Checked syllables are those ending in a glottal stop or a glottalized coda, and these often develop into glottalized phonation, as seems to have occurred with the 'tense' phonation in Yi languages (Kuang, 2011, 2012). White Hmong (Hmongic, East Asia) is known to have a lexical tone (the -*m* tone) that is variably described as checked or as creaky, which suggests that its pronunciation may vary between the two (see discussion by Ratliff, 1992, p. 12 and references therein). Because laryngealization is often derived historically from glottal stops, it is rare for glottal stops to contrast synchronically with other forms of glottalization. Exceptions are found, however, notably in the linguistic convergence areas of Mesoamerica, mainland southeast Asia, and southwestern Africa, where glottal stops are licit codas and where some tones and/or vowels can be laryngealized. In some of these languages, such as Tlacolulita Zapotec and Yucatec Maya, a phonemically laryngealized vowel may even be followed by a glottal stop (For Tlacolulita Zapotec, see the UCLA Phonetics Lab Archive.[1] For Yucatec Maya, see Frazier (2009)). Other languages with both laryngealized vowels and coda glottal stops do not allow the two to cooccur, e.g. in the Mon-Khmer language Krathing

---

[1]`http://archive.phonetics.ucla.edu/Language/ZPK/zpk.html`. (Last visited April 21, 2013).

Chong, spoken in Thailand (Silverman, 1995, §3.1.2). The fact that laryngealization and glottal stops can occur even in the same syllable in some languages suggests that the two are produced differently, though the extent to which they differ is unknown.

As already mentioned, laryngealized phonation can be associated with lexical tone in a variety of languages (see overview by Silverman (1995)). Some examples of tones with glottalization include Mandarin Tone 3 (low or dipping) and sometimes Tones 2 and 4 (rising and falling) (Davison, 1991; Belotel-Grenié & Grenié, 1997), Hanoi Vietnamese Tones B2 (falling with glottalization throughout), C1 (falling with slight laryngealization and breathiness) and C2 (falling-rising with glottalization in the middle) (Michaud, 2004; Brunelle, 2009; Brunelle et al., 2010), Cantonese Tone 4 (mid falling) (Vance, 1977; Yu, 2010), Latvian Tone 3 (Lehiste, 1972), and there is anecdotal evidence of laryngealization associated with the lowest tones in Yoruba (Yu, 2010, citing Welmers, 1973). It is not surprising that tone and laryngealized phonation may interact, given that both involve the intrinsic muscles of the larynx (see Section 2.2.2). The physiological mechanisms that are used to produce a very low F0 are likely to cause some laryngealization (Gerratt & Kreiman, 2001), and a very low F0 has similar acoustic characteristics to vocal fry (creak), a form of irregular voicing with a very low frequency (see Section 2.3.3).

Besides the cases for which laryngealization is derived from a glottal stop, the converse is also known to occur. That is, there are known diachronic cases where a laryngealized tone in a language comes to be realized with full glottal adduction, a glottal stop. For the Hanoi Vietnamese Tones B2 and C2 as well as Latvian's Tone 3, the laryngealization can be realized with complete vocal fold closure (as [ʔ]), which suggests that those tones are developing glottal stop targets from laryngealization (Lehiste, 1972; Nguyen & Edmondson, 1998; Brunelle, 2009). Indeed, Mandarin tone 3 is now often produced with full glottal closure in the Beijing variety, resulting in a [vʔv] sequence. The Mandarin case is clearly one where laryngealization developed into a glottal stop, and could explain how some languages (e.g. Santa Ana del Valle Zapotec (Esposito, 2010)) have 'rearticulated' vowels produced as a [vʔv] sequence. There is some work on Cantonese and Mandarin showing that glottalization

associated with certain tones is used by the native listener to recover the phonemic tone (Yu, 2010; Yu & Lam, 2011, but cf. results by Garellek, Keating, Esposito & Kreiman (2013) for White Hmong). This suggests that even if glottalization is a by-product of the production of certain tones in those languages, its phonetic role is not trivial.

Lastly, glottal closure gestures are also found as components of other classes of sounds that will not be studied closely in this dissertation. A summary of these sound classes and other uses is presented in Table 2.1 (p. 13).

Table 2.1: Summary of some linguistic gestures which sometimes or always involve a glottal stop gesture.

| Gesture | Explanation & References |
| --- | --- |
| Glottal stop length | Some languages contrast [ʔ] and [ʔː], e.g. Lebanese Colloquial Arabic (Ladefoged & Maddieson, 1996, p. 75), Maltese (Hume, Gett, Hovey, Scudieri & Spagnol, 2010), and North Saami (Baal, Odden & Rice, 2012). |
| Ejectives | By default involve glottal adduction (Ladefoged & Maddieson, 1996, p. 78). |
| Implosives | Glottalization is language dependent (Ladefoged & Maddieson, 1996, pp. 87–90) |
| Pharyngeals | Pharyngeal sounds may involve glottal adduction (Laufer & Condax, 1981). Glottal stops can be accompanied by pharyngeal constriction (Roach, 1979; Esling & Harris, 2005). |
| [-ATR] | Can involve glottal constriction, e.g. in Maa (Guion, Post & Payne, 2004). |
| Epiglottals | Usually involve constriction at the glottis, which is complete for the epiglottal stop [ʡ] (Esling, 1996, 2003). |
| Suprasegmental glottalization | E.g. Danish stød (Fischer-Jørgensen, 1987, 1989; Basbøll, 2003) |
| Morphological glottalization | Boundary lengthening in Finnish (Suomi, Toivanen & Ylitalo, 2008, §5.2), Bole (Gimba, 2000, pp. 164–166) and Northern Paiute (Thornes, 2003; Houser, Kataoka & Toosarvandani, 2006) |
| Pragmatic glottalization | E.g., cues to word boundaries in English (Nakatani & Dukes, 1977); cues to turn-taking in French, English, and Finnish (Malécot, 1975; Laver, 1980; Local, Wells & Sebba, 1985; Local, Kelly & Wells, 1986; Ogden, 2001) |

## 2.2 Relevant anatomical and physiological properties and linguistic models of glottalization

This section will describe key anatomical structures involved in glottalization, with the goal of understanding how glottal stops are produced. The last part of this section will focus on linguistic models of glottalization, which are based on laryngeal physiology.

The anatomical structures relevant to the study of glottal stops can be broken down into three distinct parts: the glottis and vocal folds, the subglottal structures, and the supraglottal (mostly laryngeal) structures. These three parts will be described below. The following summary is based on details from Hirose (1997) and Kreiman & Sidtis (2011), unless otherwise cited.

### 2.2.1 The larynx, vocal folds, and glottis

The larynx is suspended by muscles, ligaments, and membranes from the hyoid bone in the neck, in front of the esophagus. The larynx consists of three main cartilages: the thyroid, cricoid, and a pair of two arytenoid cartilages, shown in Figure 2.2 (p. 15).

The vocal folds (or vocal cords) are a layered structure comprising a cover (mucosa epithelium and superficial lamina propria), a transition layer (middle and deep layers of the lamina propria), and a body (the vocalis muscle, see Section 2.2.2). The vocal folds stretch from just below the notch at the front of the thyroid cartilage (the thyroid prominence) to the two arytenoid cartilages. The vocal folds range from 17-24 mm in length for men and 13-17 mm for women, but can stretch by about 3-4 mm. When the vocal folds are sufficiently abducted, a space is formed between them, called the glottis. The term subglottal refers to any structure below the glottis (i.e. the trachea and lungs), whereas supraglottal refers to anything above the glottis, which includes additional laryngeal structure as well as the vocal tract. The glottis can be divided into two parts. The space between the vocal folds (roughly two thirds of the glottis) is often referred to as the ligamental or membranous glottis, in

contrast to the space between the arytenoid cartilages, known as the cartilaginous glottis. Typically, [?] is thought to involve the closure of both of these parts of the glottis.



(a) Antero-lateral view                    (b) Posterior view
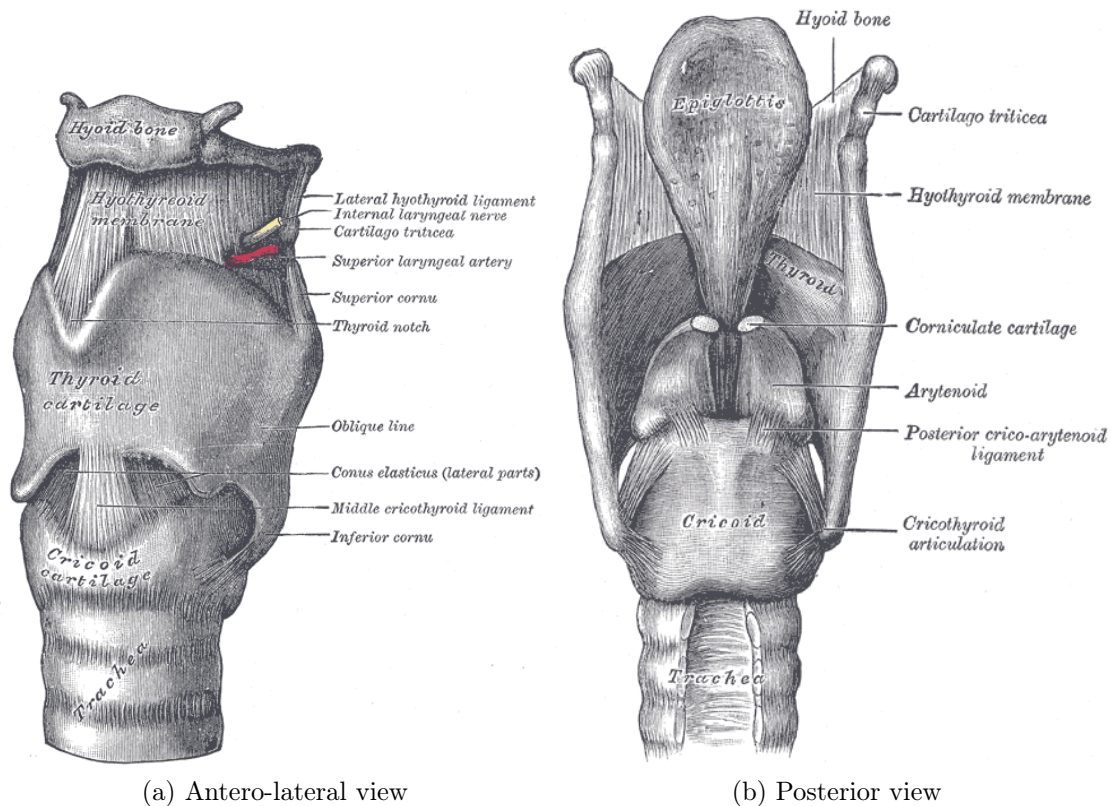
Figure 2.2: Major cartilages (thyroid, cricoid and arytenoid) of the larynx. From Gray (1918), panels 951 & 952, `http://www.bartleby.com/107/236.html`.

### 2.2.2    Intrinsic laryngeal muscles

The muscles of the larynx can be divided into intrinsic and extrinsic muscles. The extrinsic muscles are responsible for the anterior-posterior and upward-downward movements of the larynx, connect the larynx to other parts of the body, and help stabilize it. The intrinsic muscles, on the other hand, connect the different laryngeal cartilages to one another. There are five intrinsic laryngeal muscles, named after the cartilages to which they connect: posterior cricoarytenoid (PCA), lateral cricoarytenoid (LCA), interarytenoid (IA), thyroarytenoid (TA), and cricothyroid (CT). Contraction of the CT helps stretch the vocal folds by tilting

the cricoid and thyroid cartilages together, and is therefore used for raising the fundamental frequency (F0) of the voice, along with the TA. Aside from the CT, the remaining intrinsic muscles can be divided into two sub-groups: abductors, which pull apart the vocal folds; and adductors, which bring the folds together. Abduction of the vocal folds is achieved through the contraction of the PCA. Vocal fold adduction can be caused by contraction of the LCA, which closes the posterior part of the membranous glottis and the anterior part of the cartilaginous glottis, the IA (which causes adduction of the arytenoids), and/or the TA. Contraction of the TA tenses and bulges the body of the vocal folds. This contraction can help close the middle of the vocal folds. The medial part of the TA is sometimes called the vocalis (VOC) muscle, but the terms TA and VOC are often used interchangeably. The VOC is mostly thought to aid in controlling vocal fold stiffness and mass rather than adduction (Hirose, 1997). The activity of these various intrinsic laryngeal muscles has been studied for glottal stops using electromyography (Faaborg-Andersen, 1957; Hirano & Ohala, 1967; Gay et al., 1972; Lindqvist, 1972; Hirose & Gay, 1973; Fischer-Jørgensen, 1987; Ludlow, Sedory & Fujita, 1991). These studies usually show an increase in activation of the adductor intrinsic laryngeal muscles (LCA, IA, and VOC), which is expected if glottalization represents a state of increased average glottal closure over each cycle (cf. Ladefoged's continuum model of phonation types, Section 2.2.5). Because the CT is not consistently contracted during glottal stops (Lindqvist, 1972; Ludlow et al., 1991), glottal stops are not forcibly accompanied by an increase in pitch.

### 2.2.3 Phonation and vocal fold vibration

Strictly speaking, phonation is defined by the production of sound at the vocal folds, though often it is used interchangeably with voicing, which is vocal fold vibration. For example, in linguistics, non-modal phonation is often used to mean non-modal voicing. The mechanism of voicing is usually characterized by the myoelastic aerodynamic theory (van den Berg, 1958). According to this theory, the combination of tissue elasticity (e.g. altering the stiffness of the intrinsic laryngeal muscles) and aerodynamic forces is responsible for initiating, sustaining,
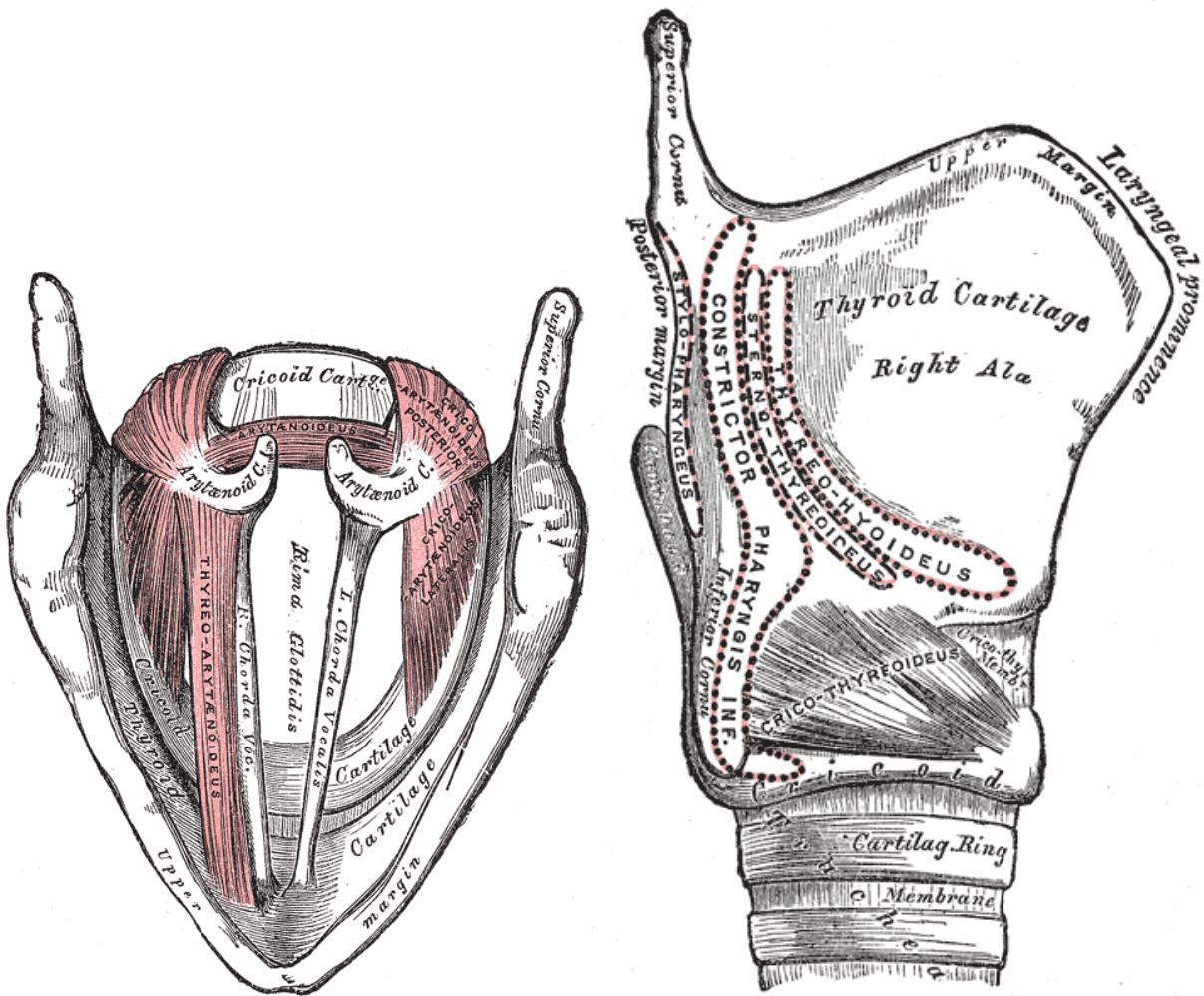
(a) Laryngoscopic view        (b) Side view

Figure 2.3: The intrinsic laryngeal muscles. The CT is visible from the side view (b). From Gray (1918, panels 960 and 957), `http://www.bartleby.com/107/illus960.html` and `http://www.bartleby.com/107/illus957.html`.

and ending the vibration of the folds. Modal vibration usually cannot start until after the vocal folds are brought together or nearly so, in order to build up a subglottal pressure of 3-5 cm $H_2O$ (Titze, 1992). Such a pressure will force the vocal folds open if they are stiff enough to produce modal voicing. Increased or decreased stiffness will require corresponding changes in the subglottal pressure in order to initiate voicing. Bringing the vocal folds together (i.e. producing a full glottal stop) will thus result in a faster phonation onset than if the vocal folds are apart, because the subglottal pressure will build up faster if no air can escape through

the glottis. Glottal stops can therefore be used to initiate voicing at phrase onsets, which might be relevant for understanding why vowel-initial words are frequently produced with vocal fold adduction. However, the possible connection between glottal stops and phonation onset has yet to be explored.

Once voicing is initiated, the natural elastic recoil of the folds and aerodynamic factors such as the Bernoulli effect and vortices in the airflow as it leaves the glottis are responsible for the glottis' closing. Reopening of the glottis is achieved in roughly the same way that vibration is initiated, but typically requires a lower minimal subglottal pressure.

The opening and closing of the vocal folds occur from bottom-to-top due to their complex structure and the direction of airflow. The vocal folds tend to close more rapidly than they open. The stiffer the vocal folds, the more abruptly they close, resulting in higher-amplitude high-frequency components in the sound spectrum.

### 2.2.4   Supraglottal structures

Above the vocal folds lie the ventricular (or false/vestibular) folds, whose anatomical and physiological properties are described in detail by Bailly (2009). The vocal folds are separated from the ventricular folds by the laryngeal ventricle (vestibule of Morgagni). The ventricular folds can adduct as well as vibrate, and their adduction, either partial or complete, is often seen during glottal stops and glottal reinforcement, e.g. in English (Fujimura & Sawashima, 1971; Roach, 1979) and Thai (Harris, 2001). It is often thought that adduction of the ventricular folds, aided by the activation of the LCA and TA muscles, helps prevent air from passing through the glottis (Hirose, 1997), thus ensuring full glottal closure.

The ventricular folds, epiglottis, and aryepiglottic folds form what is called the laryngeal vestibule (Zemlin, 1988; Edmondson, Esling, Harris & Wei, 2004), which is formed anterio-posteriorly by the epiglottis and arytenoids, and laterally by the aryepiglottic folds that connect the arytenoids to the sides of the epiglottis (see Figure 2.4, where the corniculate cartilages are superior to the arytenoid cartilages).

Trachea

Corniculate cartilage

Cuneiform cartilage

Aryepiglottic fold

·Ventricular fold

Vocal fold

Tubercle of epiglottis

Epiglottis

Vallecula

Median glosso-epiglottic fold
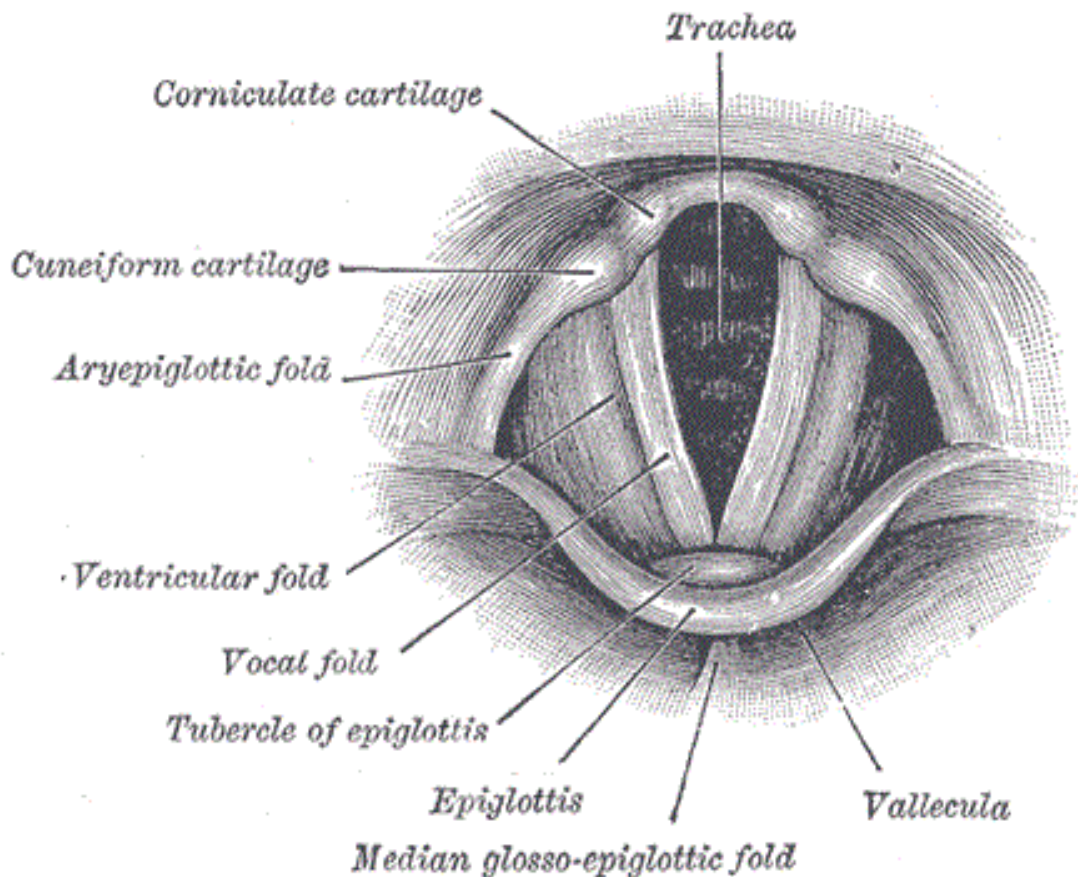
Figure 2.4: Laryngoscopic view of the interior of the larynx. Adapted from Gray (1918), panel 956, `http://www.bartleby.com/107/illus956.html`.

Upward lie the pharynx and remaining vocal tract. Contraction of the pharynx and/or epiglottal retraction can influence vocal fold movement, probably because movement in the pharynx is typically accompanied by vocal fold adduction (Esling et al., 2005).

### 2.2.5   Linguistic models of glottalization

Two models, both of which are based on laryngeal physiology, are particularly relevant for analyzing the relationship between glottal stops and other forms of glottalization. The relationship between the glottal stop and phonation types that result in glottalization (here, 'laryngealized' phonation) is clearly represented in Ladefoged's *glottal continuum* of phona-

19

tion types, which are defined in terms of the aperture between the arytenoid cartilages (Ladefoged, 1971; Gordon & Ladefoged, 2001). If laryngealized phonation (called 'creaky' in Figure 2.5) is represented by a small average glottal opening, then not surprisingly, a glottal stop is at the extreme of the continuum, as shown by 'glottal closure' in Figure 2.5.

Most open ←————————————————→ Most closed

Phonation type    Voiceless    Breathy    Modal    Creaky    Glottal closure

Figure 2.5: Ladefoged's continuum of phonation types, from Gordon & Ladefoged (2001).

However, the relationship between glottal stops and laryngealized phonation types can also be expressed using more elaborate models of phonation, e.g. the *valves model* of the throat (Esling & Harris, 2005; Edmondson & Esling, 2006). This model expresses laryngeal and supralaryngeal configurations using a system of hierarchical valves in the larynx and pharynx. The six valves represent articulatory postures, the first three being immediately relevant for glottal stops and creaky voice: Valve 1 represents glottal adduction and abduction (equivalent to Ladefoged's glottal continuum); Valve 2 represents the partial covering and damping of vocal fold vibration by the ventricular folds (what Edmondson & Esling (2006) call ventricular incursion); and Valve 3 involves compression of the arytenoids and aryepiglottic folds. Edmondson & Esling (2006) show that laryngealized voice (which they use to refer to both creaky and harsh voice) involves these first three valves, two of which are above the glottis. As mentioned earlier, supraglottal activity is often found in glottal stops. Indeed, some researchers claim that ventricular incursion is a prerequisite for full adduction of the vocal folds (Hirose, 1997; Esling et al., 2005). In their work Esling, Edmondson, Harris, and their collaborators often refer to [ʔ] as a 'moderate' glottal stop that necessarily involves some ventricular incursion, as well as some constriction of the whole laryngeal vestibule (Edmondson et al., 2004). Roach (1979) also found, for his own speech, that glottal stops involved some constriction of the laryngeal vestibule. This is in contrast to a 'strong' or 'massive' glottal stop (equivalent to the epiglottal stop, denoted by [ʡ]), which in addition

to glottal adduction involves strong pharyngeal/aryepiglottic activity (Esling, 1996; Esling et al., 2005).

According to both models, laryngealization and glottal stops may be closely tied articulatorily. The continuum model puts a heavier emphasis on glottal stops being the extreme form of glottal closure, and therefore they should be expected to occur in environments with less lenition, e.g. phrase-initially. Conversely, if a glottal stop were lenited, the continuum model would predict that it would be identical to laryngealized phonation (with varying degrees of average glottal closure). In contrast, the valves model does not state whether glottal stops represent a more extreme articulation than laryngealization, and in fact they can be rendered stronger through epiglottic involvement. An incomplete glottal stop could therefore be realized in ways differing from laryngealized phonation in the valves model.

## 2.3 Previous studies of glottal stops and glottalization

This section provides an overview of the relevant literature on glottal stops, and is divided into three parts corresponding to the three overarching questions of this dissertation: (1) how are glottal stops produced? (2) how are they perceived? (3) where do they occur, and why?

### 2.3.1 Production studies

What we know about glottal stop articulation is derived from two main sources: phonetic studies of languages with segmental glottal stops, and speech science studies of voicing onset ('attack') and offset. Articulatory studies of glottal stops and glottal gestures make use of several methods of investigation: electromyography (EMG) of the intrinsic laryngeal muscles responsible for vocal fold closure (see Section 2.2.2), electroglottography (EGG) to analyze the voicing around [?], and/or high-speed imaging of the vocal folds and surrounding structures to visualize laryngeal movement around [?]. In this section, I will discuss studies

that employ methods which I use in this dissertation: EGG and direct visual observation (imaging) of the larynx.

### 2.3.1.1  EGG studies

A useful method of studying glottalization is electroglottography (EGG), which reflects the degree of contact between the vocal folds. When vocal fold contact is highest, viz. during glottal closure, the EGG signal is able to pass from one electrode to the other with the smallest electrical resistance. Relative contact is usually measured using the parameter commonly known as 'contact quotient,' the approximate proportion of the glottal cycle during which there is contact (CQ) (Rothenberg & Mashie, 1988). How contact is defined and measured differs across studies, as described in detail by Baken & Orlikoff (2000, ch. 10), Henrich, d'Alessandro, Doval & Castellengo (2004), and Herbst & Ternström (2006). Differences are mostly based on definitions of the opening and closing instants in a glottal cycle. In particular, because a precise time point for the opening instant is usually hard to define, the opening instant is often defined by a threshold, whose value differs by study (see Herbst & Ternström (2006) for more details).

EGG measures are typically used to characterize different types of phonation. For example, laryngealized phonation often exhibits higher values of CQ across languages in which it is contrastive, for example in the recent work by Keating, Esposito, Garellek, Khan & Kuang (2011), Esposito (2012), and Kuang & Keating (2012). Because it is typically used as a tool for measuring aspects of voicing, EGG is not normally used for describing (voiceless) glottal stops in the linguistic literature. But some studies have focused on the phonation around /ʔ/. For example, Esposito & Scarborough (2004) found slightly higher CQ in Pima vowels preceding /ʔ/, and even during some instances of /ʔ/ that were lenited to laryngealized phonation. Nonetheless, CQ did not vary much for vowels preceding /ʔ/ when compared with modal vowels.

Of course, there are other measures that can be made from the EGG pulses or their derivatives, including measures of pulse skewness, speed of the opening and contact phases, and measures of the entire pulse shape using functional data analysis. (For good reviews of EGG measures, see Baken & Orlikoff (2000) and Kuang & Keating (2012)). For example, Esposito & Scarborough (2004) found that phonemic laryngealization had EGG peak velocity values similar to modal vowels before /ʔ/ in Santa Ana del Valle Zapotec, indicating that modal voice around glottal stops is similar in voice quality to laryngealization.

Thus, EGG is a useful method for measuring *voicing* around glottal stops. Previous work, though limited, suggests that glottalization will result in higher measures of EGG contact quotient, as well as other measures like peak velocity. CQ though relates more closely to glottal stops than velocity measures, because it is a measure of relative contact during voicing, and captures the distinction between phonemic modal vs. laryngealized voices well.

### 2.3.1.2 Imaging studies

Many researchers in speech science and phonetics have looked at glottal stop articulation via direct visual observation of the glottis (laryngoscopy) using a fiberscope or rigid endoscope through the nasal or oral cavity. Direct imaging of the vocal folds before and during phonation onset (attack) has a long history, and it has long been noted that hard glottal attacks show more constriction than normal or breathy ones (Moore, 1938). Later studies began to quantify the results of imaging different forms of phonation onset. For example, Cooke et al. (1997) found that hard glottal attacks showed increased maximum vocal fold velocity and stiffness but a shorter duration gesture than normal or breathy attacks. Švec & Schutte (1996) used a method called *kymography* to produce functional images (kymograms) of the larynx as sequences of vocal fold images. Kymograms are useful for showing the trajectory of vocal fold movement, and are easily interpretable. For this reason, their use has become increasingly popular (e.g., in Wittenberg, Tigges, Mergell & Eysholdt, 2000; Bailly, 2009;

Orlikoff et al., 2009; Bailly, Henrich & Pelorson, 2010, and Freeman et al., 2012). In Chapter 3, I use kymograms to visualize the phasing of ventricular incursion.

Glottal stops, either as independent phones or accompanying other linguistic gestures (e.g. as glottal reinforcement) have been investigated using direct observation in many studies, e.g. by Fujimura & Sawashima (1971); Roach (1979); Zawaydeh (1999); Harris (2001); Esling & Harris (2003); Esling et al. (2005), Esling et al. (2007), and Edmondson et al. (2011). However, such studies usually do not quantify rate of occurrence of ventricular incursion, focusing instead on glottal closure duration or pharyngeal constriction. Therefore, we do not know how often ventricular incursion occurs, and little imaging work has been done to investigate and measure the phasing between ventricular incursion and glottal stop articulation.

Thus, imaging studies have shown that glottal stops or hard glottal attacks are characterized by vocal fold stiffening and supraventricular constriction. However, it is still mostly unclear how glottal stops are produced – i.e., what changes to vocal fold vibrations occur, when such changes occur, and how they might explain how voicing and sustained vocal fold closure are coproduced. Moreover, although the presence of ventricular incursion during [ʔ] production has long been noted, we still do not know precisely when it occurs, which can help explain whether and for what reasons supraglottal constriction is indeed necessary for the production of glottal stops.

### 2.3.2 Perceptual studies

The perception of glottal stops has received little attention in the phonetic literature. The first major study was by Hillenbrand & Houde (1996). In this study, the authors manipulated the F0 and/or amplitude (RMS energy) for sequences of /oʔo/ in four experiments. Listeners were native English speakers who were asked to say when they heard a glottal stop in intervocalic position. The results indicate that locally reduced amplitude and (especially) F0 are sufficient to cue the presence of a glottal stop, even when no silence was present.

In contrast to the perception of glottal stops, the perception of creaky or laryngealized phonation has received more attention in the literature (Pham, 2003; Gerfen & Baker, 2005; Brunelle, 2009; Frazier, 2009; Kuang, 2011; Yu & Lam, 2011; Kuang, 2012; Kuang & Keating, 2012; Garellek et al., 2013). Importantly, many of these show that manipulation of the cues shared by glottal stops and laryngealized phonation (notably F0) improves identification of laryngealized phonation. For example, Gerfen & Baker (2005) replicated some of the methods in Hillenbrand & Houde 1996 for laryngealized vs. modal vowels in Coatzospan Mixtec. The authors found that, in the absence of spectral or durational differences between modal and laryngealized vowels, F0 and/or amplitude dips are used by listeners to perceive a laryngealized vowel.

Therefore, it appears that manipulations of cues shared by both glottal stops and laryngealized phonation (dips in F0 and/or amplitude) result in similar changes in identification, regardless of whether participants are asked to detect glottal stops (Hillenbrand & Houde, 1996) or laryngealized phonation (Gerfen & Baker, 2005). This suggests that glottal stops might be hard to detect in creaky conditions. In Chapter 4, I test this in an English word identification task.

### 2.3.3 Acoustic studies of glottalization

A major question I will address in this dissertation is what factors are most important in predicting pre-vocalic 'inserted' glottal stop occurrence in English. In many studies, researchers have looked at the 'optionality' of glottal stop occurrence by noting the presence of irregular pitch periods (i.e., visual evidence of laryngealized voicing), because laryngealization is known to be a common outcome of a glottal stop gesture. Studies looking at the occurrence of laryngealization derived from segmental glottal stops, glottal reinforcement, and/or creak include (among others) Pierrehumbert & Talkin (1992); Pierrehumbert (1995); Dilley et al. (1996); Hagen (1997); Docherty & Foulkes (1999); Huffman (2005) for English, Batliner, Burger, Johne & Kießling (1993); Kohler (1994); Hagen (1997); Pompino-Marschall & Żygis (2011) for German, Malécot (1975) for French, Blankenship (1997) for Navajo and Tagalog,

25

Jongenburger & van Heuven (1991) for Dutch, Huber (1988) for Swedish, Priestly (1976) for Slovene, and Frazier (2009) for Yucatec Maya. These studies have found that glottal stop realization is very variable (ranging from full stops to laryngealization), and that a variety of factors can predict when (non-phonemic 'optional') glottal stops occur. In what follows, I summarize some of these findings, focusing on the prevocalic, optional glottal stops in English, but occasionally making reference to similar findings for coda glottal stops and glottalization in other languages.

### 2.3.3.1 Predicting non-contrastive glottalization in the acoustic signal

Although glottalization is known to be highly variable even within-speaker, many acoustic studies (mostly on English) have found that the phenomenon is likelier to occur in certain conditions. The factors that affect the rate of occurrence of glottal stops and other forms of glottalization may be segmental, lexical, prosodic, or sociolinguistic in nature. Predicting glottal stops using quantitative measures (to assess degree instead of rate of glottalization) is still rarely undertaken (cf. Seid, Yegnanarayana & Rajendran (2012) for glottal stops in Amharic).

**Segmental factors.** Glottal stops in vowel-initial words in English were found to be more common (when phrase-internal) after a word ending in a vowel (Dilley et al., 1996), consistent with studies showing increased rates of glottalization at a vowel-vowel hiatus (Umeda, 1978; Pierrehumbert & Frisch, 1997). Umeda (1978) found that glottal stops were more likely to occur before back vowels than before front ones. Similarly, glottal stop insertion is also said to be more common before lower vowels in German, perhaps due to perceptual reasons and/or pharyngeal involvement (Brunner & Żygis, 2011; Pompino-Marschall & Żygis, 2011). Glottal reinforcement of coda stops in English is known to occur after vowels and approximants, though its occurrence is said to be optional in such environments (Cohn, 1993; Huffman, 2005). In British English, coda-/p/ has been found to be glottalized more often than /t/ or /k/ (Watt & Milroy,

26

1999), but this apparently differs across dialect (Wells, 1982). For American English, Pierrehumbert (1995) and Huffman (2005) found that glottal reinforcement was rare for coda-/t/ when it was followed by a stop, but it was common before sonorants. Coda-/p/ could be glottalized before nasals, and coda-/k/ did not undergo glottalization.

**Lexical factors.** It has been noted that in English, less frequent vowel-initial words are more likely to be produced with a glottal stop (Umeda, 1978). The same study also found glottalization to be rare on vowel-initial function words, presumably due to the close juncture between such words and the lexical items they modify, though possibly also due to their high frequency. Vowel-initial content words in German were also found to glottalize more frequently than function words (Pompino-Marschall & Żygis, 2011). Similarly, English words with a low relative frequency (frequency of the word/sum of the frequencies of that word and its phonological neighbors) showed increased temporal and spatial coarticulation of glottalized coda-stops (Garellek, 2011).

**Prosodic factors.** Prosodic factors are now known to be strong predictors of glottalization, especially in English, as this topic has been given substantial study since the Tones and Break Indices (ToBI) framework for labeling prosody was established (Silverman, Beckman, Pitrelli, Ostendorf, Wightman, Price, Pierrehumbert & Hirschberg, 1992). For example, vowel-initial words are found to have increased glottalization rates under nuclear stress in English (Pierrehumbert, 1995; Dilley et al., 1996) and in German (Pompino-Marschall & Żygis, 2011), which could explain why content words glottalize more than function words. Vowel-initial glottalization in English is known to be more common after larger phrasal boundaries (Pierrehumbert & Talkin, 1992; Dilley et al., 1996). Rate of speech (possibly due to changes in phrasing) has also been shown to affect glottalization rates for German vowel-initial words, with less frequent glottalization at faster speech rates (Pompino-Marschall & Żygis, 2011).

In her dissertation, Epstein (2002) included quantitative measures for determining the degree of 'tense' phonation in a phrase. Her results showed that both prominent (accented) and phrase-initial words displayed tenser phonation, though the precise timing of the initial tense phonation was not explored.

**Sociolinguistic factors.** Like presumably all 'optional' phonetic events, glottalization rates are known to vary according to certain sociolinguistic factors. For example, higher rates of glottalization and/or creak in American English have been found for women (Byrd, 1994; Dilley et al., 1996; Podesva & Lee, 2010; Yuasa, 2010), though the opposite was found for male speakers of two varieties of British English, suggesting that there is probably variability in glottalization rates across sex and dialects (Henton & Bladon, 1988). French and Swedish women were also found to have higher rates of glottalization than men (Malécot, 1975; Huber, 1988). Social ambition and class effects have been reported for glottalization of coda-stops in British English (Mathisen, 1999; Mees & Collins, 1999). Age is also known to affect glottalization rates in Newcastle English (Docherty & Foulkes, 1999).

In sum, a variety of factors have been shown to favor glottalization, but these factors are rarely compared one to the other, in order to determine which of them are the best predictors.

### 2.3.3.2 Descriptive assessments of glottalization from the acoustic signal

Early studies of glottal stop insertion rates were more phonological in nature, focusing on where glottal stops may occur (based on the authors' impressions) and what grammatical rules could be used to account for their insertion (Higginbottom, 1964; Roach, 1973). Since then, most work on glottal stop insertion rates has relied on acoustic cues to the segment. But because glottal stops are commonly realized with incomplete closure, various studies have devised visual landmarks for concluding that a sound is laryngealized, and then count

this as a form of glottal stop (Huber, 1988; Batliner et al., 1993; Redi & Shattuck-Hufnagel, 2001). Some studies created automated algorithms for correct classification of glottalization (Batliner et al., 1993; Surana, 2006; Seid et al., 2012). The classification schema called *Münchner Schema für Laryngalisierungs-Identification (MÜSLI)* by Batliner et al. (1993) coded glottalization/creak along six acoustic dimensions, specifying the degree of damping and number of pitch periods in the signal, and amplitude and F0 excursions. Redi & Shattuck-Hufnagel (2001) narrowed down the acoustic dimensions to four distinct cases:

1. *Aperiodicity*: "successive pitch periods for which incremental changes in duration were discontinuous. For example, a region might display a jump from a relatively shorter pitch period to a longer one, then back to a short period, or it might exhibit the converse– a sequence of long-short-long periods."

2. *Creak*: "gradual widening in pitch period, resulting in a very low fundamental frequency with associated strong damping of pitch periods." Also frequently called *vocal fry*.

3. *Diplophonia*: "regions of alternation of pitch periods (in a simple repeating pattern or a more complex pattern) which had different amplitudes, shapes, or period lengths." Also called *period doubling*.

4. *Glottal squeak*: "an instantaneous increase in fundamental frequency which was subsequently sustained for multiple periods," usually accompanied by low amplitude.

These four types of irregular pitch periods appear to be sufficient for describing glottalization as it occurs in non-pathological voice (cf. van Rossum, van As-Brooks, Hilgers & Roozen (2009) for pathological voice). Note that none of these criteria above identifies glottal stops made with complete vocal fold adduction except if they cause coarticulatory laryngealized voicing (though the frequency with which [ʔ] is accompanied by laryngealized voicing is unknown).

### 2.3.3.3 Quantitative assessments of glottalization from the acoustic signal

Several studies have used acoustic measures to quantify laryngealized phonation and how its acoustic signal might differ from normal, modal phonation. These studies are a useful accompaniment to qualitative work on glottalization, because laryngealization varies somewhat continuously in degree (cf. Ladefoged's continuum model of phonation types in Section 2.2.5), and therefore it is possible that qualitative studies overlook subtler forms of laryngealization within modal voice. This point is made explicitly by Epstein (2002) as motivation for using quantitative measures of laryngealization (in her case, via inverse filtering and LF-model fitting), though she still had to disregard the more extreme laryngealized tokens that were not fittable.

Common measures of laryngealized phonation include spectral measures, e.g. the amplitude difference in the first and second harmonics (H1-H2, or when corrected for formants, H1*-H2*), the amplitude difference in the first harmonic and the harmonic nearest the first, second, or third formants (H1-A1 or H1*-A1*, H1-A2 or H1*-A2*, and H1-A3 or H1*-A3*), noise/aperiodicity measures like cepstral peak prominence (CPP) or harmonics-to-noise ratios specified over the entire spectral domain or ranges of frequencies (HNR), measures of the first formant bandwidth (B1), and measures of variability (e.g. measures of jitter/shimmer). The spectral measures are thought to correlate with various articulatory features of laryngealization: H1-H2 with open quotient (OQ) or contact quotient from EGG (Holmberg, Hillman, Perkell, Guiod & Goldman, 1995); H1-A1/2/3 with a posterior opening of the glottis and/or the simultaneity of ligamental closure (Stevens, 1977; Hanson, Stevens, Kuo, Chen & Slifka, 2001); and H1-A1 or B1 in particular with posterior glottal opening (Hanson et al., 2001). For more details of acoustic measures of laryngealized voice quality and discussion of their presumed articulatory bases, see Gordon & Ladefoged (2001); Shue, Shattuck-Hufnagel, Iseli, Jun, Veilleux & Alwan (2010b), and Garellek & Keating (2011). Some recent studies that have used such measures to describe laryngealized phonation include Blankenship (2002) for Mazatec, Chong, and Mpi, DiCanio (2009) for Chong, Miller (2007) for Ju|'hoansi, Wayland & Jongman (2003) for Khmer, Garellek & Keating (2011) for Mazatec, Andruski (2006) for

Green Mong, Garellek (2010) for English, Korean, and White Hmong, Kuang (2011, 2012) for Yi, and Esposito (2010) for Zapotec, among others. Generally, it has been found that laryngealized phonation shows lower values of the spectral measures (H1-H2, H1-A1/2/3) than modal phonation due to an increase in higher frequency energy during laryngealization (presumably due to abrupt closure of the vocal folds), lower values of B1, and lower values of CPP/HNR due to increased aperiodicity. In addition, laryngealized phonation has often been found to have lower overall energy, likely because the vocal folds are held close together and remain closed for a large portion of the glottal cycle (Fischer-Jørgensen, 1987; Gordon & Ladefoged, 2001). However, not all the studies have found significant differences between modal and laryngealized voice, or in the expected directions, which suggests that different types or degrees of laryngealization exist across languages and/or speakers, or that the different ways of calculating these measures yield different results. Because these are measures of voicing, they do not provide any information on glottal stops produced with sustained vocal fold adduction, but they are useful for studying the effects that a glottal stop has on the surrounding voicing and how long such effects might last.

In sum, there exist many studies of the acoustics of laryngealized phonation and laryngealization thought to derive from glottal stops, and such studies have found many factors that influence where glottal stops (usually incomplete ones) may occur. However, there is still little understanding of when glottal stops are realized with full closure, and whether their predictors differ in a systematic way from those that predict other forms of glottalization. This is relevant for understanding whether complete glottal stops inserted vowel-initially are extreme forms of glottalization, or whether they are predicted by different (lexical, segmental) factors.

## 2.4   Summary of background

In the previous sections, I surveyed the literature on glottal stops and glottal-stop-like gestures in languages of the world, provided an overview of the relevant anatomy involved in

glottal stop production, and reviewed linguistic models of glottal stop articulation. Despite an abundance of studies of glottal stops, it is still unclear whether they are really glottal (vs. supraglottal or glottoventricular), if glottal stops are perceived distinctly from creak, and why they are so common before word-initial vowels. In the following chapters, I will report the results of three studies in which I aim to solve this lack of understanding.

# CHAPTER 3

# Are glottal stops really glottal?

## 3.1 Introduction

In the International Phonetic Alphabet (IPA), a glottal stop [ʔ] is characterized as a voiceless plosive produced at the glottis (though, for other definitions of glottal stops in the phonetic literature, see Esling et al. (2005, §1.2)). Thus, as with most IPA sounds, [ʔ] is simply defined with respect to its place and manner of articulation and with no indication of the primary articulator. However, the primary articulator of glottal stops is assumed to be the vocal folds, which approximate and close the glottis.

Nevertheless, direct observation of the larynx during the production of [ʔ] has revealed that glottal stops often – if not always – involve movement of more than just the vocal folds. Other supraglottal (mostly epilaryngeal) structures that may be involved include the arytenoids and ventricular folds (which adduct) and the laryngeal vestibule (which narrows). The laryngeal vestibule is the inferior portion of the larynx and comprises the epiglottis and aryepiglottic folds (Zemlin, 1988; Esling, 1996, 1999; Harris, 1999, 2001; Esling et al., 2005). The narrowing of the laryngeal vestibule is characterized by retraction of the epiglottis and forward movement of the arytenoids via the contraction of the aryepiglottic folds. A schematic of the relevant structures (laryngoscopic view) is shown in reproduced in Figure 3.1.

Based on their research on glottal stops, Esling and colleagues posit a continuum of glottal stop productions, which can be symbolized as a continuum from [ʔ] to [ʔ]. The weakest – yet still complete – glottal stops are characterized by closure of the vocal folds
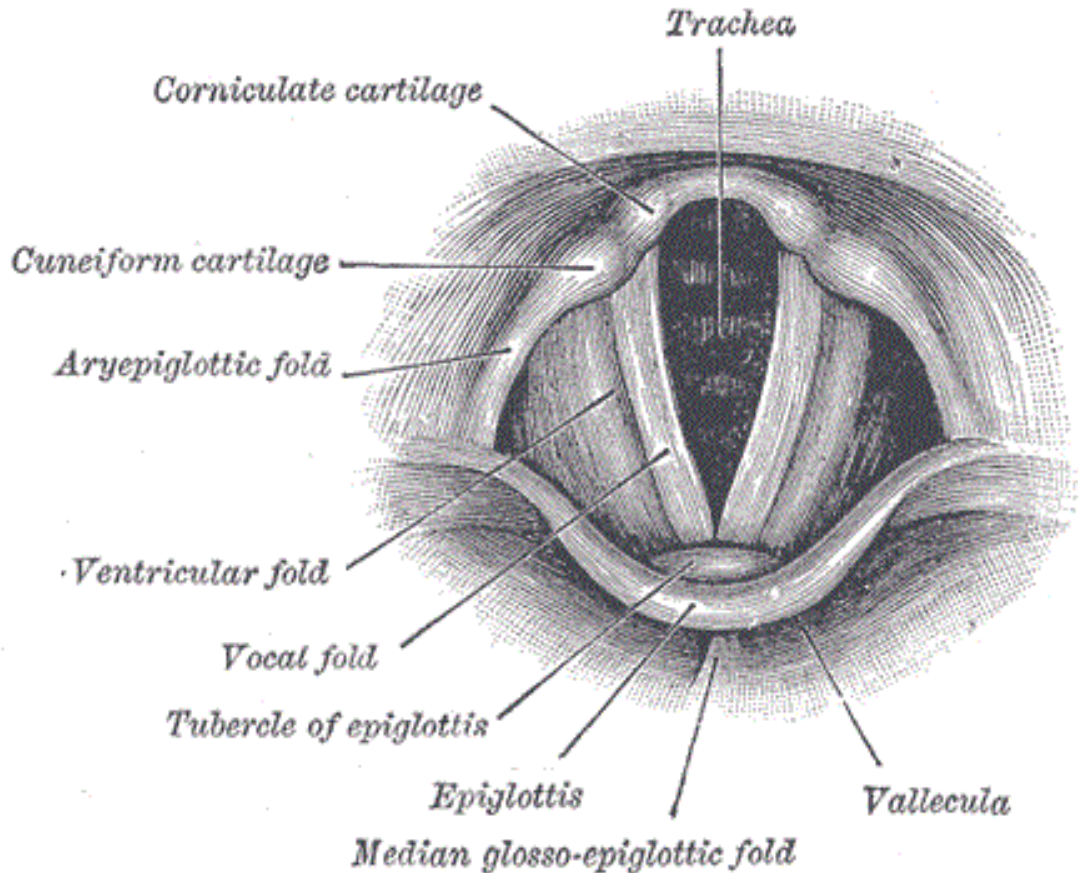
Figure 3.1: Laryngoscopic view of the interior of the larynx, with anterior part at the top of the image. Adapted from Gray (1918, panel 956), `http://www.bartleby.com/107/illus956.html`. The same figure appears in Chapter 2 (Figure 2.4). In this study, the focus will be on the vocal and ventricular fold movements.

and slight approximation (incursion) of the ventricular folds; moderate glottal stops involve more ventricular incursion, and the more extreme (epi)glottal stops involve full ventricular incursion, constriction of the laryngeal vestibule, and even pharyngeal constriction (Harris, 2001; Carlson & Esling, 2003; Esling & Harris, 2003; Esling, 2003; Esling & Harris, 2005; Esling et al., 2005; Edmondson & Esling, 2006; Esling et al., 2007; Hassan & Esling, 2007; Moisik & Esling, 2011). Note that according to this glottal stop continuum model, [ʔ] always involves supraglottal structures, because even weak glottal stops have some ventricular incursion. For this reason, Esling et al. (2007) state (p. 585) that 'a "glottal stop" cannot be uniquely glottal but is at least glottoventricular.' This is in contrast to Ladefoged's

continuum model of glottal strictures (Ladefoged, 1971; Gordon & Ladefoged, 2001), where [ʔ] is defined solely by the degree of closure of the glottis.
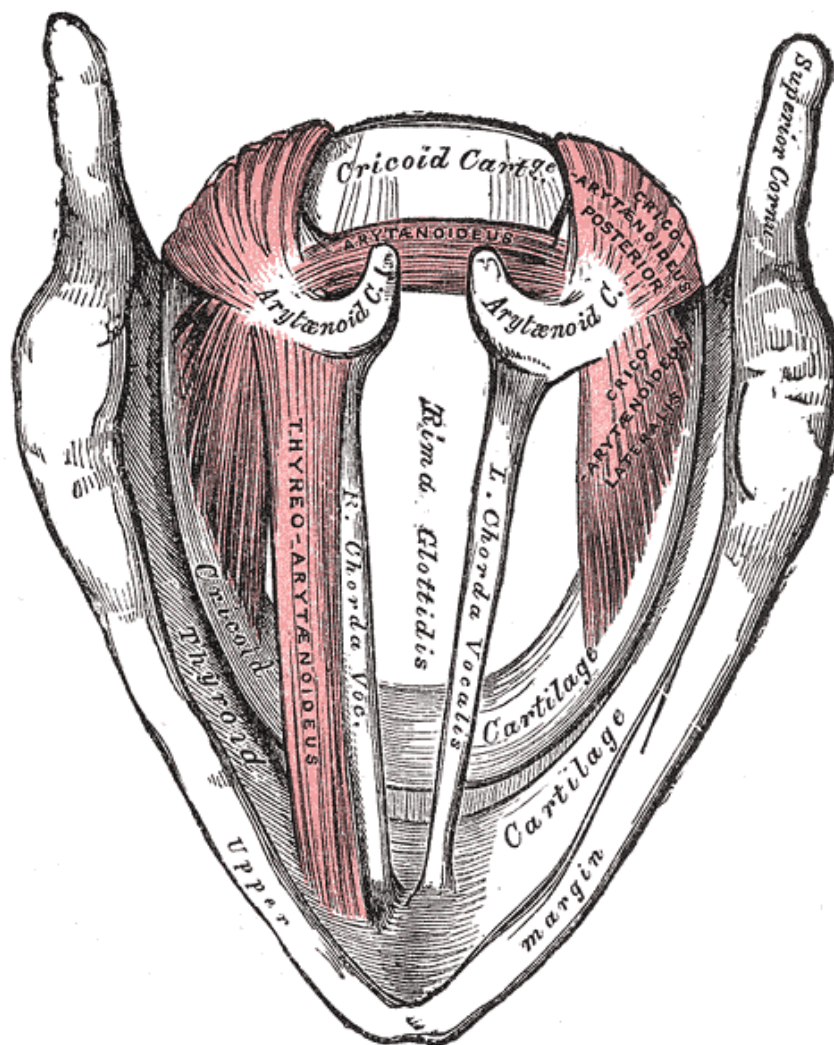
Figure 3.2: Laryngoscopic view of the intrinsic laryngeal muscles, excluding the cricothyroid. From Gray (1918, panel 960), `http://www.bartleby.com/107/illus960.html`. For view of CT, see Figure 2.3 in the Introduction (p. 17).

There are several reasons – either incidental or deliberate – for which supraglottal (epilaryngeal) articulations would be involved in the production of glottal stops. Supraglottal articulation may be incidental because the intrinsic laryngeal muscles involved in vocal fold adduction are attached to supraglottal structures. Thus, supraglottal articulation during glottal stops may be a by-product of muscular coupling between the glottal and supraglottal

articulators. For example, although the precise details of the ventricular folds' musculature are still controversial, most researchers recognize that constriction of the thyroarytenoid (TA, see Figure 3.2) can induce ventricular incursion (Reidenbach, 1998; Sakakibara, Kimura, Imagawa, Niimi & Tayama, 2004; Esling et al., 2007; Bailly, 2009), and the same muscle is involved in glottal stop production (Hirano & Ohala, 1967; Ludlow et al., 1991; Cooke et al., 1997).

But supraglottal involvement during [ʔ] articulation may also be deliberate: it helps to either close the glottis, or sustain closure, or both close the glottis and sustain the closure. Because Esling and colleagues have found that most instances of [ʔ] are 'moderately' articulated (with at least partial ventricular approximation), Esling et al. (2005) claim that 'a glottal stop typically requires supraglottic reinforcement to arrest the vibration of the vocal folds' (p. 402). Yet if supraglottal articulators aid in the arresting of vocal fold vibration, then this implies (1) that glottal stops which are not preceded by voicing (e.g., Utterance-initial [ʔ]) should not have supraglottal involvement, and (2) that supraglottal involvement should be visible and likely strongest at the moment of glottal closure.

Although there is much laryngoscopic evidence for supraglottal articulation during the production of glottal stops, findings from other types of studies are less conclusive. Acoustically, supraglottal (especially pharyngeal) constriction should result in formant transitions (e.g., F1 raising for pharyngealization (Alwan, 1989; Stevens, 2000)) into and out of glottal stops, yet such transitions are not typically found (Ladefoged & Maddieson, 1996). Modeling studies have recently shown that ventricular incursion is not necessary for the production of glottal stops (Moisik & Esling, 2012). The discrepancy between findings from imaging studies on the one hand and acoustic/modeling studies on the other can be due to different methodologies and/or to a lack of understanding of the acoustic/physiological implications of glottalization. Imaging studies that rely on trans-oral or -nasal endoscopy are highly invasive, allowing only for small corpora (from few speakers) and hyper-articulated speech. These conditions could easily increase the odds of recording hyper-articulated glottal stops, which would indeed have supraglottal constrictions. But it is also possible that supraglot-

36

tal articulations – especially light pharyngealization or epilaryngeal constriction – do not perturb the acoustic output enough to be easily detected. Until we know more about the acoustic consequences of epilaryngeal articulations, we cannot conclude with certainty that most glottal stops show no acoustic evidence of supraglottal articulations.

The goal of this chapter is to answer key phonetic questions regarding how glottal stops are produced: whether glottal stops are always produced with supraglottal articulation, and if supraglottal articulation occurs, at what point during [ʔ] articulation. By 'supraglottal articulation,' I will refer specifically to the ventricular folds, though ventricular movement is often accompanied by supraventricular movement as well. To test whether glottal stops are always produced with ventricular incursion, I conducted a production experiment using trans-oral laryngoscopy to analyze glottal closure for [ʔ], whether ventricular folds are present during glottal stops, and where ventricular incursion occurs with respect to glottal closure.

## 3.2    Method

The high-speed imaging was done using trans-oral laryngoscopy in the Department of Head and Neck Surgery at the UCLA Medical School. The scoping was done by a trained practitioner (either Dr. Gerratt or Dr. Chhetri of the Department of Head and Neck Surgery).

### 3.2.1    Stimuli

Because the trans-oral laryngoscope is placed above the tongue, participants could only produce vowels and glottal sounds. Moreover, only /i/-like vowels were elicited, because the high-front tongue position moves the epiglottis anteriorly, which allows for a better view of the glottis. In practice though, the /i/ tends to sound closer to [ɨ] because the scope prevents the tongue from raising and fronting as far as usual.

There were four groups of stimuli: the first group consisted of [ʔi] sequences. Participants were instructed to say three instances of /ʔi/ at three pitch heights (low-mid-high), resulting

in [ʔì, ʔī, ʔí]. The second group of stimuli consisted of [iʔ] sequences at three pitch heights: [ìʔ, īʔ, íʔ]. The initial vowels usually began with some breathiness ([ʰiʔ]).

To obtain different phrasings, the third and fourth groups consisted of [ʔi] sequences in reiterant speech. For the third group, speakers were instructed to say the arithmetic statement "1+1+1+1", with each syllable replaced by [ʔi]. This results in the phrase [('ʔi) (ʔi 'ʔi) (ʔi 'ʔi) (ʔi 'ʔi)], where parentheses indicate boundaries of smaller intonational units (e.g., intermediate phrases), and stress marking indicates expected phrasal prominence. For the fourth group of stimuli, speakers were instructed to say the arithmetic statement "(1+1)+(1+1)", with each syllable replaced by [ʔi]. This results in the phrase [('ʔi ʔi 'ʔi) ('ʔi) ('ʔi ʔi 'ʔi)].

For each group of stimuli, participants were told to 'produce glottal stops,' but were not explicitly instructed to produce full glottal stops (as opposed to laryngealization).

### 3.2.2  Participants

Five phonetically-trained native English speakers (mean age = 31.2, two women and three men) participated in the tasks.

### 3.2.3  Task and high-speed videoendoscopy

The first two groups of stimuli were recited in succession; the third and fourth groups were recited separately. Each utterance lasted no more than six seconds. For each utterance, high-speed images of the vocal folds were recorded using a Phantom V210 camera (Vision Research, Wayne, NJ) at a sampling rate of 10,000 frames per second, with a resolution of 208x352 pixels. The camera was mounted on a Glidecam Camcrane 200 (Glidecam Industries, Kingston, MA). The A/D converter of the camera (Module 9223, National Instruments, Austin, TX) had a voltage resolution of 16 bits with input range ±10V. Voice signals were synchronously recorded with a Brüel & Kjær microphone (diameter = 1.27 cm; type 4193-L-004), held approximately 7 cm from the corner of the speaker's mouth, and were directly digitized at a sampling rate of 60 kHz (conditioning amplifier: NEXUS 2690, Brüel & Kjær,

Denmark; bandpass filtering of microphone signal between 20 Hz and 22.4 kHz; analog-to-digital converter: voltage resolution 16 bits, input range ±5V). However, the audio recordings will not be analyzed below.

In total, 100 glottal stop tokens were recorded (20 per speaker: 3 in the onset group, 3 in the coda group, and 7 in each phrasal groups). However, 13 of these (1 in the onset group, 8 in the first phrasal group, and 4 in the second phrasal group) were excluded because the image was obscured, either because the camera was not well positioned, or because the epiglottis blocked the view of the glottis. Thus, in total 87 tokens are analyzed below.

### 3.2.4    Analysis

The glottal areas of the complete utterances were visualized using *GlotAnTools*, a software toolkit supplied by the Department for Phoniatrics and Pedaudiology of the University Hospital, Erlangen, Germany. For each glottal stop token, the time points of three articulatory events were identified: [ʔ] closure (here defined as the cessation of voicing), [ʔ] end (onset of voicing), and the point at which ventricular incursion reached its maximum. For the first onset [ʔ] in a [ʔi], the moment of [ʔ] closure is undefined, because glottal closure is defined here with respect to voicing. Similarly, the last glottal closure duration for the final [ʔ] in sequences of [iʔ] was undefined, because there was no subsequent voicing.

In addition to the identifying crucial time points, I also noted if glottal stop closure and ventricular incursion were complete. Complete glottal stop closure is defined as the absence of a membranous or cartilaginous glottis during the glottal stop. Presence of a posterior gap or incomplete vocal fold adduction was thus noted for each token. For most cases of [ʔ] produced with a posterior gap, the arytenoids gradually closed fully during the period of glottal 'closure,' so that by the time the glottal stop was released, the posterior gap had disappeared. On the other hand, for instances of [ʔ] produced with incomplete *vocal fold* closure (which a priori had not been expected to occur), the anterior gap remained open

throughout the glottal stop's duration. Examples of full glottal stops produced with different closure types are shown in Figure 3.3.



(a) complete closure          (b) posterior gap          (c) anterior gap
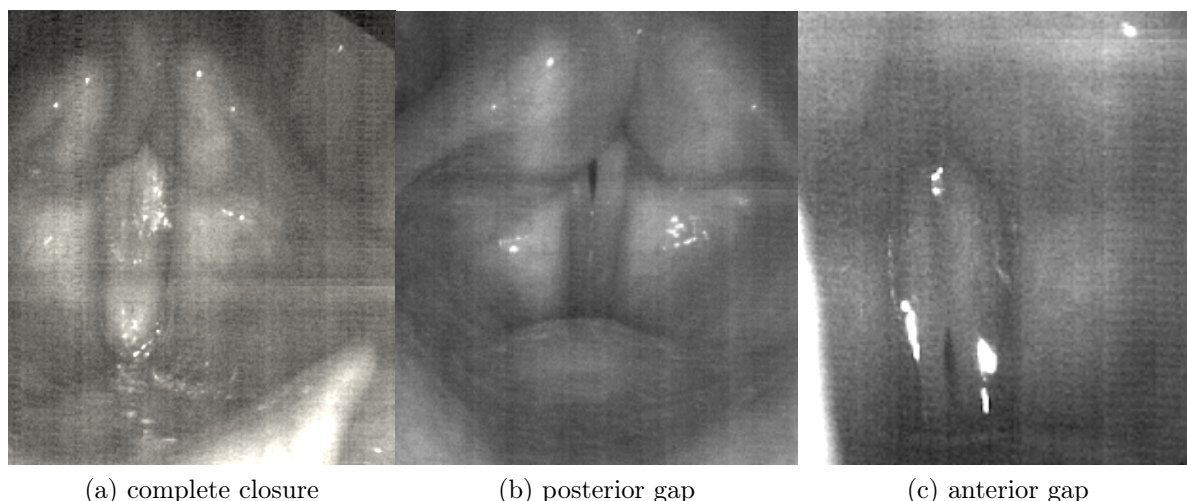
Figure 3.3: Examples of glottal stop closure types for cases of full [?].

Types of ventricular incursion were labeled as either 'no incursion,' 'incomplete incursion,' or 'full incursion.' If there was no incursion, then the ventricular folds did not medialize. (For three tokens, there was some ventricular movement at the moment of glottal closure, presumably due to muscular coupling from the stiffening of the vocal folds. However, the ventricular folds did not medialize afterwards, so these were counted as cases of no incursion). Incomplete ventricular incursion meant that the ventricular folds medialized but did not reach complete adduction (i.e., they did not touch one another). Full incursion implied that the ventricular folds medialized to the point of touching one another. Examples of full glottal stops produced with different types of ventricular incursion are shown in Figure 3.4, and a summary of qualitative labels is shown in Table 3.1. Ventricular incursion was usually accompanied by constriction of the aryepiglottic folds. However, such constriction was not analyzed in this study.

| (a) no incursion | (b) partial incursion | (c) full incursion |
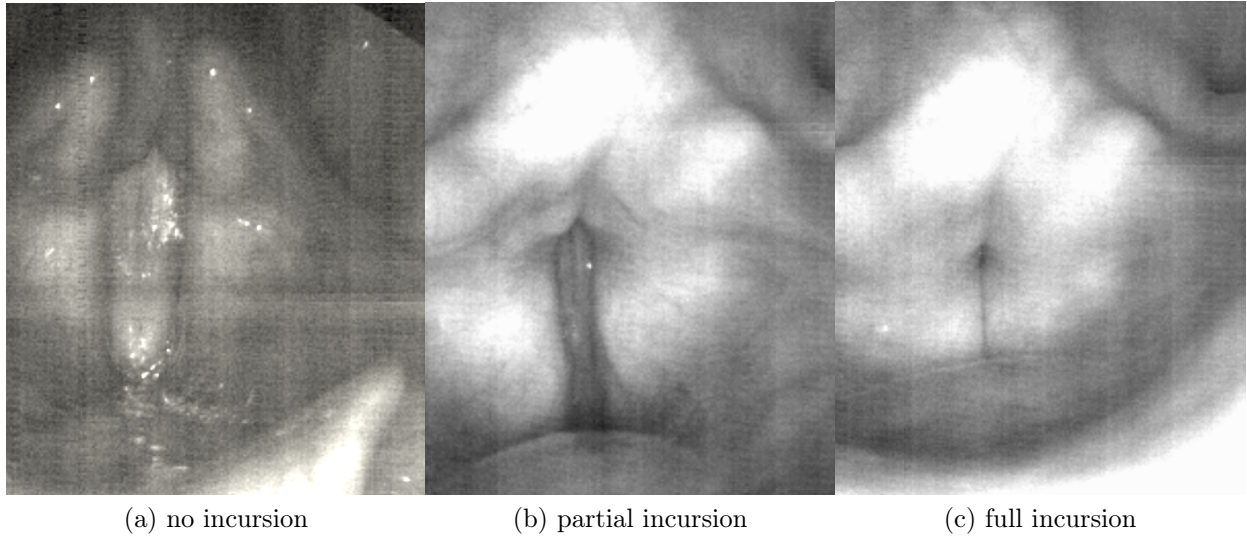
Figure 3.4: Examples of ventricular incursion types at maximal constriction points.

Table 3.1: Qualitative labels for vocal and ventricular fold adduction during glottal stop targets.

| Major label | Minor label | Explanation |
| --- | --- | --- |
| | No stop | No cessation of voicing (laryngealization) |
| Incomplete /ʔ/ | Posterior gap | Incomplete inter-arytenoid closure, but no voicing |
| | Anterior gap | Incomplete vocal fold closure, but no voicing |
| Full /ʔ/ | | Full closure of both membranous and cartilaginous glottises |
| No incursion | | No medialization of ventricular folds |
| Incomplete incursion | | Medialization but ventricular folds do not touch |
| Full incursion | | Ventricular folds touch one another |

The maximal and minimal points of ventricular incursion were analyzed using kymography (Švec & Schutte, 1996; Wittenberg et al., 2000). Laryngoscopic kymograms are produced from a line of pixels that is extracted at the same location for a sequence of frames. In the kymograms shown below, a bottom-to-top displacement refers to a horizontal displacement on a standard laryngoscopic view of the larynx. An example is shown in Figure 3.5 (p. 42). The kymogram (on top) is a sequence of frames corresponding to a horizontal pixel line over a laryngoscopic view of the glottis. The line is placed horizontally on the image at the location of maximal ventricular fold adduction. The red dashed line in the kymogram represents a line of pixels for a single frame during ventricular incursion, shown at the bottom left panel

41
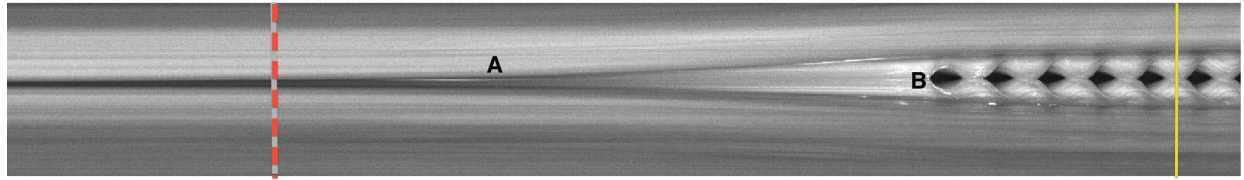
Time (frame)

(a) Kymogram
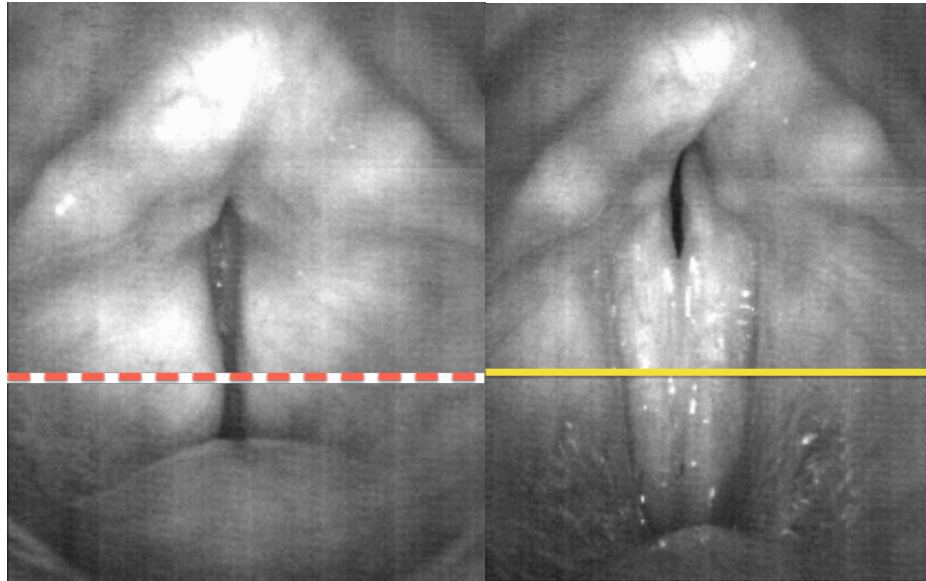


(b) Glottal stop closure  (c) Vocal fold closure during voicing

Figure 3.5: Example of a kymogram (a). The red dashed line in the kymogram represents the line of pixels during ventricular incursion (b), whereas the yellow line in the kymogram represents the line of pixels during glottal closure during voicing (c). 'A' is the approximate frame at which the ventricular folds begin abducting, and 'B' the approximate onset of voicing. Notice that the vocal folds stay closed long after the ventricular folds begin abducting (i.e., point B is later than point A).

of the figure. The yellow solid line in the kymogram represents a line of pixels for a single frame during voicing, shown at the bottom right panel of the figure.

During voicing, the opening and closing of the vocal folds is visible from the periodic dark (vocal folds open) and lighter (vocal folds closed) images, as can be seen on the rightmost portions of Figure 3.5. When the ventricular folds are fully medialized (as in the beginning of the kymogram in Figure 3.5), then the image is gray with a black line in the center, corresponding to where the ventricular folds are nearly touching. Kymograms are thus

42

useful for visualizing (for onset glottal stops) whether the ventricular folds have completely abducted at the onset of voicing, and (for coda glottal stops) where incursion begins. In other words, they can be used as a method of inferring the points during glottal stop production at which the ventricular folds begin and end their incursion (see samples in Appendix A). This in turn can be used as evidence that the ventricular folds are involved in making a glottal stop in coda position (rather than just reinforcing one).

## 3.3    Results

In the results section, first I analyze all the data from the four groups combined, without regard to glottal stop phrasing or syllable position. Subsequently, I subset the data according to the glottal stop position conditions (onset vs. coda) and the two phrasing conditions.

### 3.3.1    Overall results

The qualitative results for type of glottal stop closure are shown in Figure 3.6. Ninety percent of the glottal stops were produced as stops, rather than as laryngealization. This is expected, because participants were told to produce glottal stops, and because the speech style was very formal. Most target glottal stops (69%) were produced as full stops with no membranous or posterior gap. Fourteen percent of target productions were glottal stops with incomplete closure of the cartilaginous glottis (i.e., with a posterior gap). Interestingly, a smaller percentage of glottal stops (7%) were characterized by cessation of vibration with incomplete closure of the membranous glottis (i.e., with an anterior gap). In other words, cessation of vocal fold vibration for the production of [ʔ] can occur without complete glottal closure, though there were no documented cases of a glottal stop produced with incomplete closure of *both* membranous and cartilaginous glottises.

The distribution of glottal stop types is divided by speaker in Figure 3.7. Two participants (M1, F1) only produced glottal stops with full closure. The other three participants produced glottal stops with other articulations. Only one participant (M3) produced all four types of

glottal stops, and this speaker was also the only one to produce glottal stops with anterior gaps (stops with incomplete vocal fold closure).

Next, the glottal stops are divided into the types of ventricular incursion that occurred during their production. The results for all participants groups together are shown in Figure 3.8a (p. 46). Only five percent of tokens had no ventricular incursion, whereas the majority of tokens (78%) had incomplete incursion. The within-subject results are shown in Figure 3.8b (p. 46). One participant (M3) only produced glottal stops with incomplete incursion, whereas two participants (M2 and F2) produced glottal stops with all three types of ventricular incursion.
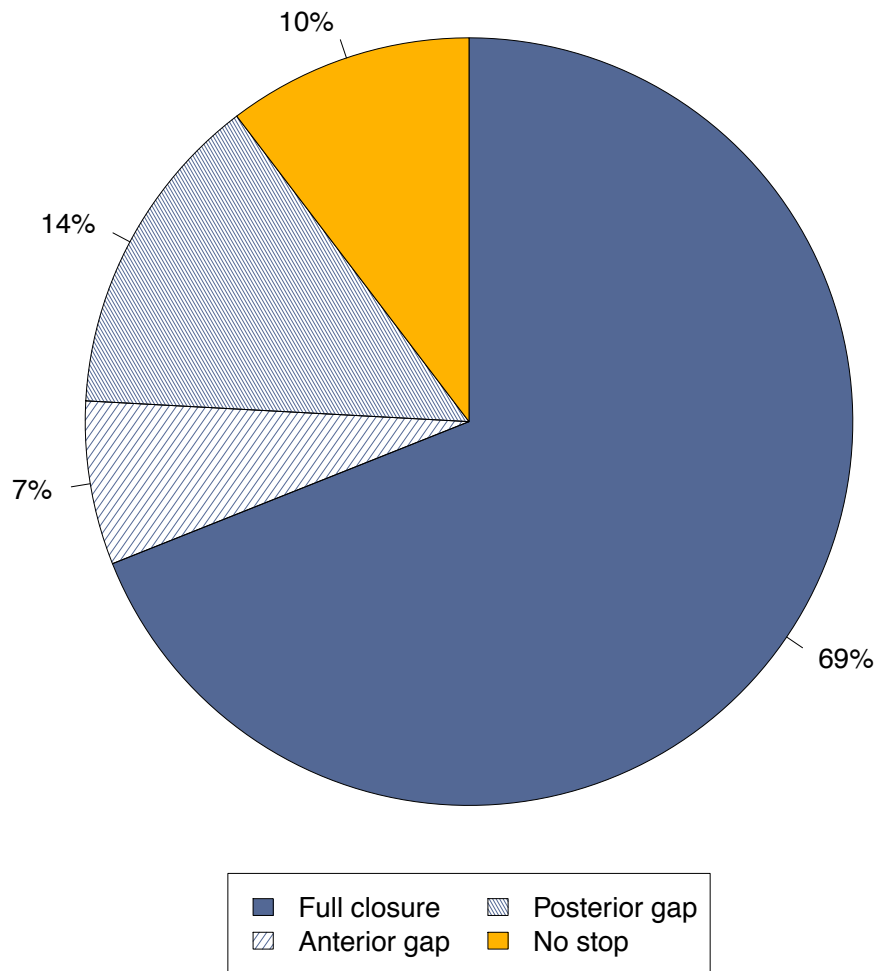


Figure 3.6: Distribution of /ʔ/ closure types in full dataset.
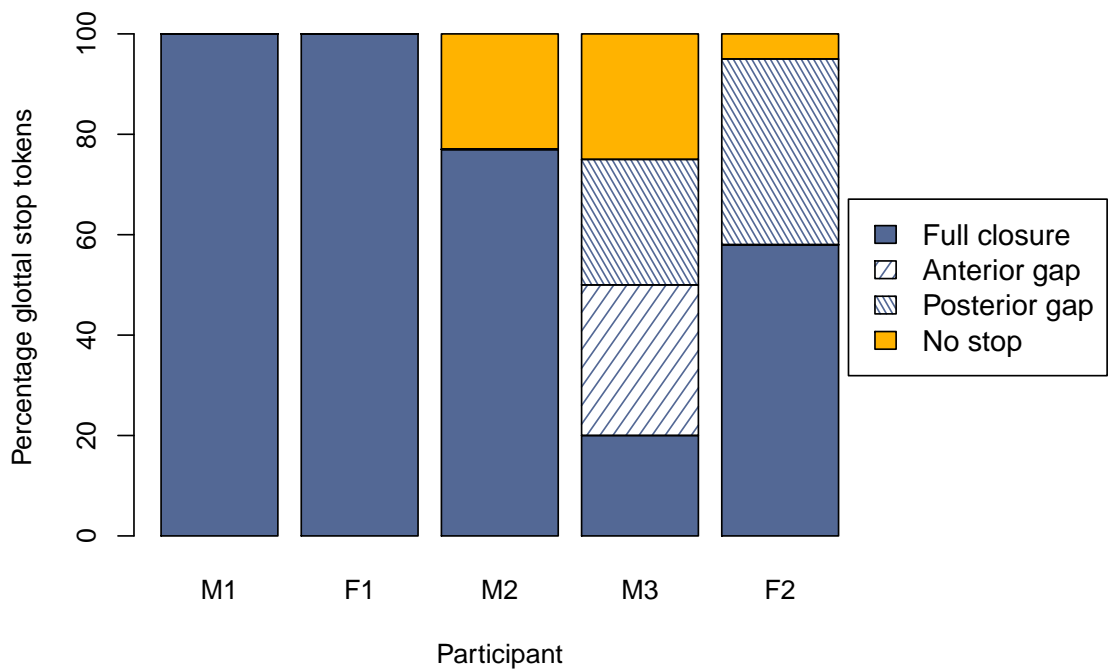
Figure 3.7: Distribution of /ʔ/ closure types by speaker.
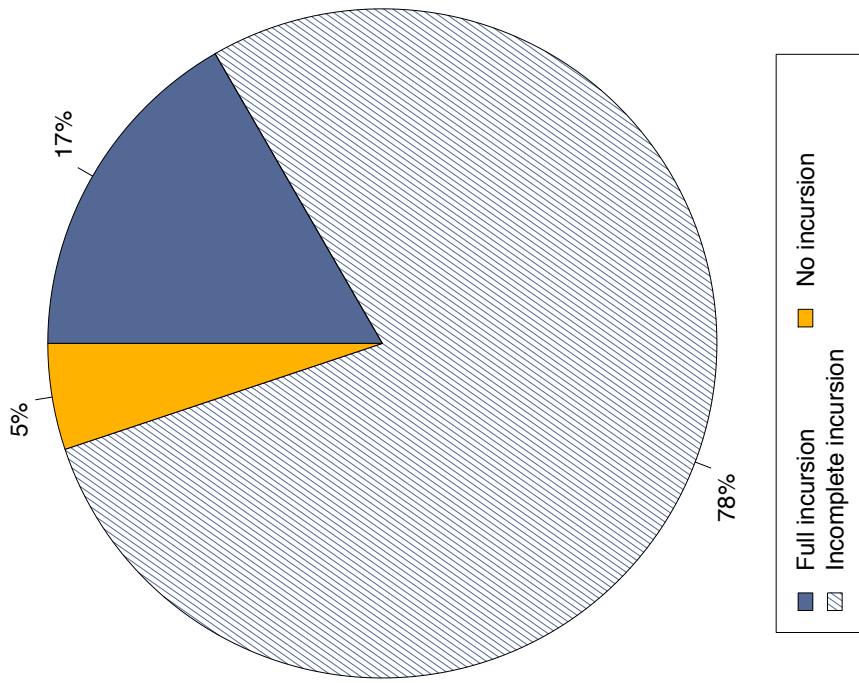
(a) Distribution of ventricular incursion types during full [?]
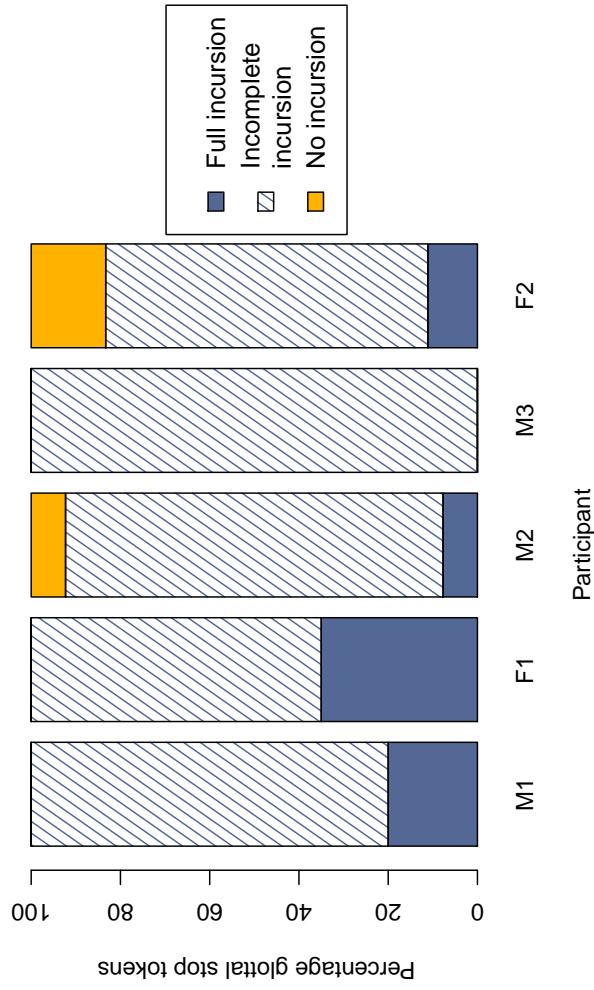


(b) Distribution of ventricular incursion types by speaker

Figure 3.8: Distribution of ventricular incursion types overall and by speaker in full dataset.

The distribution of ventricular incursion type according to the type of glottal stop produced is shown in Figure 3.9. Interestingly, all three types of ventricular incursion were found for [ʔ] produced with adduction of both the vocal folds and the arytenoids, which means that ventricular incursion is not associated only with the more extreme glottal stops. Note that for the tokens of [ʔ] produced with a posterior gap and with full ventricular incursion, the posterior gap closed before incursion reached its maximum (but after vocal fold adduction). Glottal stops produced incompletely (as a form of laryngealization) all showed some amount of ventricular incursion, as has been reported for creaky voice (Edmondson & Esling, 2006).
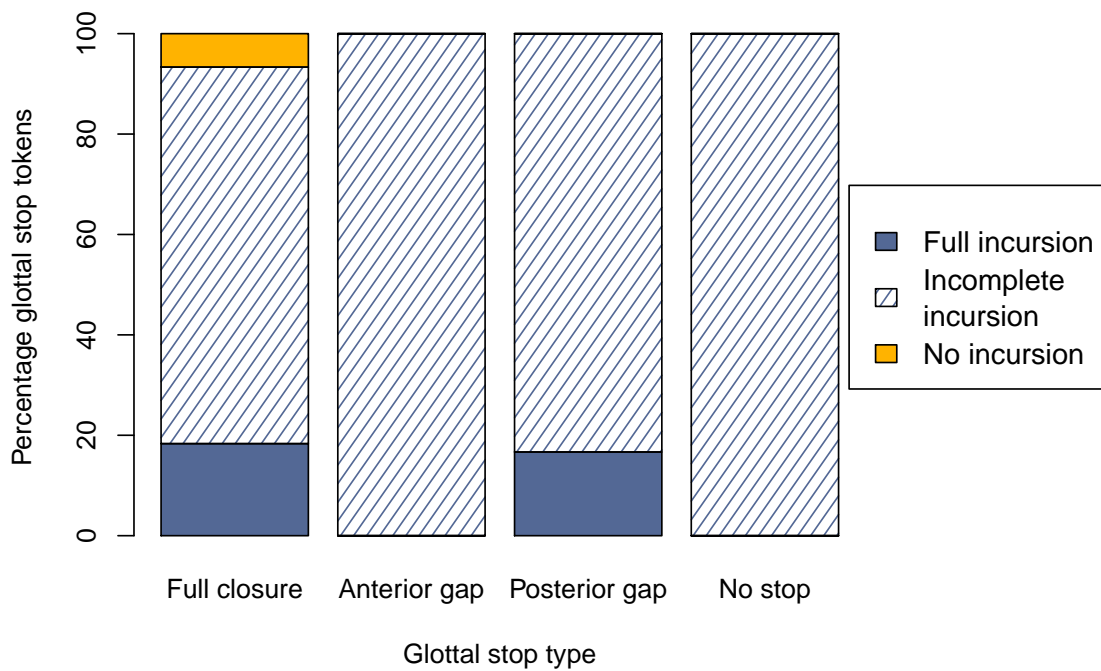


Figure 3.9: Distribution of ventricular incursion types by glottal stop closure in full dataset.

The average [ʔ] duration was 185.5 ms. For glottal stops with ventricular incursion, the ventricular folds reached their maximal degree of incursion on average 93.9 ms (about 50.6%) into the [ʔ] closure. Thus, ventricular incursion is strongest in the middle of the glottal stop.

### 3.3.2 Effect of glottal stop position

In this section, only cases of [ʔ] in the onset (15 tokens) or coda (14 tokens) conditions are investigated. The distributions of glottal stop and ventricular incursion types are shown in Figure 3.10. There were no instances of anterior gaps for the glottal stops in the onset or coda conditions (only in the phrasal ones). Although the data are limited ($N$=14 for onsets in group 1 ([ʔi]), 15 for codas in group 2 ([iʔ]), there is a tendency for coda glottal stops to be more 'strongly' produced. That is, [ʔ] in coda position is likelier to have full glottal closure (Figure 3.10, left panel) and full ventricular incursion (Figure 3.10, right panel). Furthermore, onset [ʔ] rarely has full ventricular incursion.



(a) Glottal stop types

(b) Types of ventricular incursion
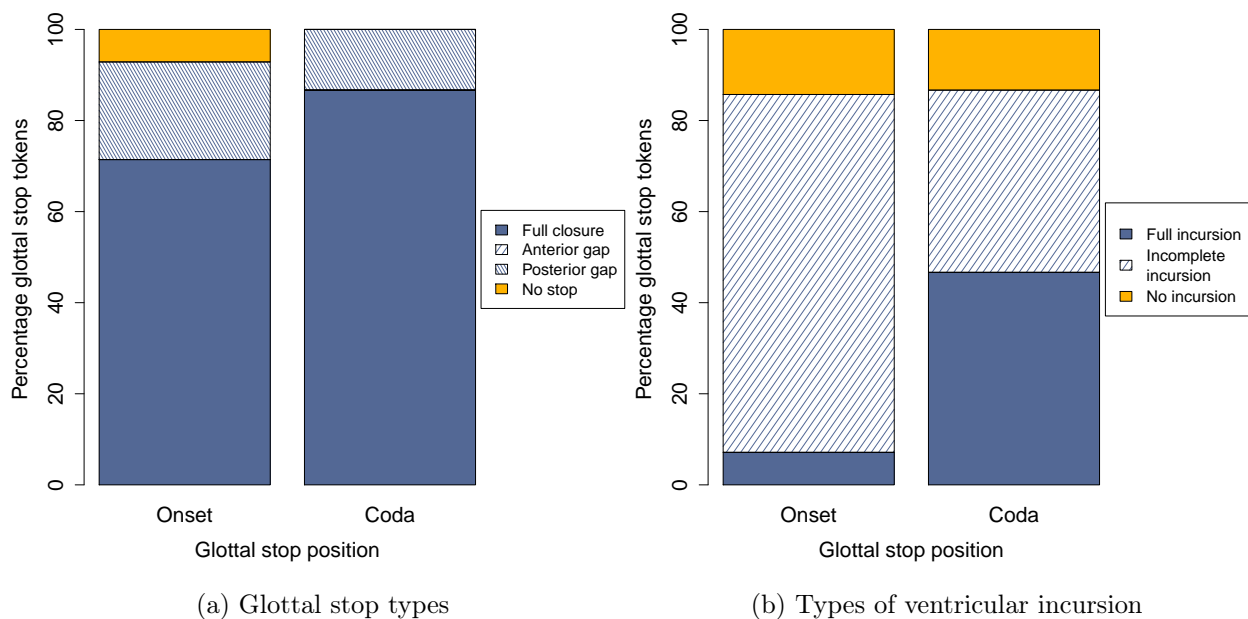
Figure 3.10: Effect of [ʔ] position on the type of glottal stop closure (a) and ventricular incursion (b) for stimuli groups 1 [ʔi] (onset) and 2 [iʔ] (coda).

For onset glottal stops, the point of minimal ventricular incursion relative to the onset of voicing was assessed kymographically (see samples in Appendix A). For each kymogram, a horizontal line was drawn at the top ventricular fold's medial edge after the ventricular folds had fully abducted. For onset glottal stops, the approximate location of full ventricular ab-

duction was taken to be the point in the kymogram at which the top ventricular fold's medial edge reached the line of maximal ventricular abduction (marked by an arrow in Figure 3.11, top panel on p. 49). For the purpose of this study, I noted whether ventricular abduction occurred before/after phonation onset, which is defined by the onset of regular opening and closing phases of the vocal folds. If ventricular abduction occurred after phonation onset, I also noted at which pulse (numbered from the onset of voicing) it occurred.

For coda glottal stops, a line was drawn at the top ventricular fold's medial edge before the ventricular folds began their incursion (marked by an arrow in Figure 3.11, bottom panel). The approximate location of the start of ventricular incursion was taken to be the point in the kymogram at which the top ventricular fold's medial edge went below the line of maximal ventricular abduction (marked by an arrow in Figure 3.11, bottom panel).
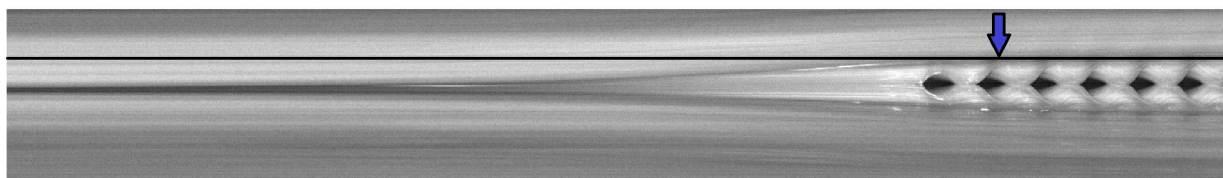


(a) Onset position
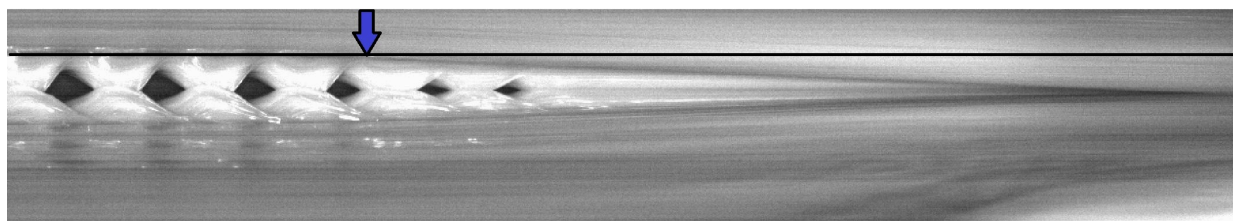


(b) Coda position

Figure 3.11: Examples of the kymographic analysis. The maximal ventricular abduction is noted using a horizontal line. For the onset [ʔ] example (a), the ventricular folds have fully abducted during the second pulse, with the approximate time marked by the arrow. For the example of a coda [ʔ] (b), ventricular incursion begins two pulses prior to the glottal stop onset.

For onset glottal stops, the ventricular folds had fully abducted on average 1.7 pulses after the onset of voicing (e.g., Figure A.1b in Appendix A), but half showed some ventricular incursion at the onset of voicing, whereas half showed fully abducted ventricular folds at the

time when voicing began (e.g., Figure A.1a). On the other hand, 77% of coda glottal stops showed ventricular incursion starting before full glottal closure, on average 2.5 pulses before glottal closure (e.g., Figure A.1d in Appendix A). Note that in this study the moment of glottal stop closure is taken to be the cessation of voicing. This provides additional support that coda [ʔ] is more strongly articulated than onset [ʔ]: coda cases are more likely to show ventricular incursion during voicing, and ventricular incursion is present for more glottal pulses in coda position compared with onset position.

### 3.3.3 Effect of phrasing

Now we turn to the other half of the data, which are divided into two subsets. For the stimuli in groups 3 and 4 (i.e., those produced in phrases), the descriptive statistics for the effect of phrasing on [ʔ] closure and duration, as well as ventricular incursion, are shown in Table 3.2. Not surprisingly, cases of incomplete glottal stops are more common in phrase-medial and unstressed positions. Cases of full ventricular incursion tend to be more common at the start of the utterance and/or in strong positions (phrase-initial and stressed). Glottal stops also tend to be shorter and weaker in phrase-final (though stressed) and in unstressed positions.

Table 3.2: Statistics for /ʔ/ by location in phrase for stimuli in groups 3 & 4.

(a) Phrase 1: (1+1+1+1)

|  | [('ʔi) | (ʔi | 'ʔi) | (ʔi | 'ʔi) | (ʔi | 'ʔi)] |
|---|---|---|---|---|---|---|---|
| Full closure (%) | 100 | 50 | 50 | 75 | 50 | 50 | 50 |
| Anterior gap (%) | 0 | 25 | 25 | 0 | 0 | 25 | 25 |
| Posterior gap (%) | 0 | 25 | 25 | 25 | 25 | 25 | 25 |
| No closure (%) | 0 | 0 | 0 | 0 | 25 | 0 | 0 |
| Full incursion (%) | 33 | 25 | 25 | 0 | 0 | 0 | 25 |
| Incomplete incursion (%) | 67 | 75 | 75 | 100 | 100 | 100 | 75 |
| No incursion (%) | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| [ʔ] duration (ms) | 323 | 173 | 128 | 145 | 144 | 151 | 108 |
| % Maximal incursion | 43 | 63 | 59 | 67 | 54 | 57 | 59 |

(b) Phrase 2: (1+1)+(1+1)

|  | [('ʔi | ʔi | 'ʔi) | ('ʔi) | ('ʔi | ʔi | 'ʔi)] |
|---|---|---|---|---|---|---|---|
| Full closure (%) | 75 | 60 | 40 | 80 | 100 | 75 | 50 |
| Anterior gap (%) | 25 | 0 | 0 | 20 | 0 | 0 | 0 |
| Posterior gap (%) | 0 | 0 | 20 | 0 | 0 | 0 | 0 |
| No closure (%) | 0 | 40 | 40 | 0 | 0 | 25 | 50 |
| Full incursion (%) | 25 | 0 | 0 | 0 | 0 | 0 | 0 |
| Incomplete incursion (%) | 75 | 100 | 100 | 100 | 100 | 100 | 100 |
| No incursion (%) | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| [ʔ] duration (ms) | 118 | 60 | 79 | 150 | 178 | 110 | 105 |
| % Maximal incursion | 39 | 39 | 52 | 49 | 62 | 50 | 45 |

## 3.4  Discussion

Using high-speed videoendoscopy, several important findings regarding glottal stops were made. In terms of glottal articulation, [ʔ] may be produced via adduction of the vocal folds, the arytenoids, or both. In this study, instances of [ʔ] with closure of both the membranous and cartilaginous glottis were the most common, but of the instances of glottal stops produced as full stops, about 23% were produced with incomplete closure of either the arytenoids or the vocal folds. Thus, the whole glottis need not be closed while the vocal folds are not vibrating. Crucially, [ʔ] should therefore be characterized as glottal adduction with absence of vocal fold vibrations, rather than by the complete glottal closure. Full glottal stops produced with incomplete closure of the glottis are likely to have some noise compo-

nent during the closure, because air may pass through the glottal opening. However, because such openings are typically small (and some do not last for the whole period of glottal approximation), it is unlikely that listeners can hear these instances of glottal stops as being 'breathy.'

Another important finding concerns the role of the supraglottal articulators, in particular the ventricular folds, during glottal stop production. Although the vast majority of full [ʔ] in this study were produced with some ventricular incursion, not all instances of glottal stops had movement of the ventricular folds. The glottal stops produced in groups 3 and 4 (the phrasal conditions) almost never had full ventricular incursion (see Table 3.2). Thus, for some cases of full [ʔ] in this study – notably those produced with full adduction of both the vocal folds and the arytenoids – the ventricular folds are not required to arrest the vibration of the vocal folds (cf. Esling et al. (2005)), and not all glottal stops need be labelled as at least 'glottoventricular' instead of glottal (cf. Esling et al. (2007)). Ventricular incursion is thus common during, but not necessary for, glottal closure.

In this study, I also investigated when ventricular incursion occurred, and when (during glottal closure) the ventricular folds reach their maximal degree of constriction. The results indicate that ventricular incursion for onset glottal stops tends to end either before phonation onset, or after the first glottal pulse following [ʔ]. For coda [ʔ], ventricular incursion tends to start roughly 2.5 pulses before glottal closure, but some cases show incursion starting after glottal closure. When ventricular incursion does occur, it peaks roughly halfway during glottal closure. This supports the finding that ventricular involvement is not needed to produce a glottal stop, because one already exists by the time that the ventricular folds adduct. Therefore, three different findings support the idea that glottal stops need not have supraglottal articulation: even the vocal folds need not be entirely shut during [ʔ], some instances of full glottal stops have no ventricular incursion, and when ventricular incursion occurs, it peaks after glottal closure.

The glottal stops in this study were 'hyperarticulated,' in the sense that they were uttered in short phrases by phoneticians, and under artificial and uncomfortable circumstances.

Thus, it is especially notable that even hyperarticulated instances of glottal stops may involve no ventricular incursion. This in turn implies that in more conversational speech, glottal stops with no ventricular incursion are even likelier to be found. But if ventricular incursion is not necessary for the production of glottal stops, why does it occur in the first place? I argue that it is used to reinforce glottal closure. Speakers were instructed to produce glottal stops, and thus their production target was full closure. It is possible that once the vocal folds close, the buildup in subglottal pressure – which can force the folds open again – triggers a response in the proprioceptive receptors of the folds that signals that the folds will reopen, all else equal. If the speaker aims to sustain closure, additional constriction will therefore be needed to counteract the buildup in pressure below the glottis. Therefore, the vocal folds can be thought of as part of a hierarchical coordinative structure whose goal is to maintain the absence of voicing. If glottal constriction is not sufficient (likely because of high subglottal pressure), then additional structures 'pitch in' in a hierarchical fashion: first the ventricular folds adduct, followed by (presumably) the aryepiglottic folds and then the pharyngeal wall. (Note that movement of the supraventricular structures were not analyzed here). This could explain why the ventricular folds maximally constrict during glottal closure – rather than before or at the moment of closure. Moreover, this could help explain why glottal stops are so often realized as incomplete glottal stops in casual speech; if incomplete glottal stops are still identified as /ʔ/ (see Hillenbrand & Houde (1996) and the following chapter), then sustaining glottal closure via the recruitment of additional structures is not necessary.

Unfortunately, the phrasing conditions in this study do not provide much evidence for or against the idea that ventricular incursion is tied to subglottal pressure fluctuations; every case of /ʔ/ in the phrases was produced with some degree of ventricular incursion. Therefore, either the phrases were too short or too artificial, or the speakers aimed to produce too strongly articulated glottal stops, to determine whether ventricular incursion is more common when the subglottal pressure is high.

Lastly, the results from the study reveal a difference in production of glottal stops depending on whether they are in onset vs. coda position. When a glottal stop occurs in coda

position, [ʔ] is more likely to be more 'strongly' articulated (that is, produced with full glottal closure and/or with full ventricular incursion). This result can be due to the fact that participants were English speakers, and thus were probably more accustomed to producing [ʔ] in word-initial position than in codas. Nevertheless, a similar discrepancy between onset and coda 'glottal stops' has been found in Thai (Harris, 2001), and so it is also possible that glottal stops require additional effort to be produced in coda position because they require cessation of voicing. This is interesting, because it implies that more constriction is needed to arrest vocal fold vibrations fully than to restart voicing after a period of glottal closure. (This might seem intuitive, but it is also be possible that the reverse should be true: a tighter glottal constriction is needed in onset position in order to build up enough pressure to initiate voicing). Assuming that speakers do not always hyperarticulate /ʔ/, then glottal stops in codas, without the extra help, are more likely to be realized as incomplete stops in casual speech (a widely-accepted fact), and they are also likely diachronically to change into contrastive laryngealized voice quality or to delete entirely.

## 3.5    Conclusion

The goals of this chapter were to determine whether ventricular incursion always occurs during glottal stop production, and – if so – at what point during [ʔ] articulation does it occur. The high-speed imaging data showed that not all glottal stops are produced with ventricular incursion, and that even glottal stops produced with full closure of the vocal folds can be strictly glottal. Furthermore, ventricular incursion tends to be incomplete and to peak in constriction about halfway through the duration of the glottal stop. Thus, this study does not support the claim that ventricular incursion is necessary for the production of glottal stops, but it does support the hypothesis that incursion occurs during glottal closure as a form of closure reinforcement.

# CHAPTER 4

# Perception of glottal stops and creak

## 4.1 Introduction

The goal of this chapter is to determine the influence of phrasal creak on the perception of allophonic glottal stops in English. There have been few perceptual studies of glottal stops since the first major study by Hillenbrand & Houde (1996). In that study, the authors manipulated the F0 and/or amplitude (RMS energy) for sequences of /oʔo/ in a series of four experiments. They found that amplitude and (especially) F0 dips are used by phonetically-trained English listeners to identify glottal stops. The findings were largely confirmed in a study by Pierrehumbert & Frisch (1997), who found that a sudden drop in F0 cues intervocalic glottalization in synthesized sequences of words like 'heavy oak.' Since then, however, studies of glottalization are still rare (but see studies by van Rossum et al. (2009) and Brunner & Żygis (2011) for perception of glottalization in disordered speech and glottalization's role in the perception of vowel quality).

In contrast to the perception of glottal stops, the perception of creaky or laryngealized phonation has received more attention in the literature (Pham, 2003; Gerfen & Baker, 2005; Brunelle, 2009; Frazier, 2009; Kuang, 2011; Yu & Lam, 2011; Kuang, 2012; Kuang & Keating, 2012; Garellek et al., 2013). Importantly, many of these studies show that manipulation of the cues shared by glottal stops and laryngealized phonation (notably F0) improves identification of laryngealized phonation. For example, Gerfen & Baker (2005) replicated some of the methods in Hillenbrand & Houde (1996) for laryngealized vs. modal vowels in Coatzospan Mixtec. The authors found that, in the absence of spectral or durational differences between

modal and laryngealized vowels, F0 and/or amplitude dips are used by listeners to perceive a laryngealized vowel.

Therefore, it appears that the manipulation of cues shared by both glottal stops and laryngealized phonation (dips in F0 and/or amplitude) results in similar changes in identification, regardless of whether participants are asked to detect glottal stops (Hillenbrand & Houde, 1996) or laryngealized phonation (Gerfen & Baker, 2005). This suggests that glottal stops might be hard to detect in creaky or laryngealized conditions. Note that, as mentioned in Section 2.1.2, glottal stops can only rarely occur adjacent to phonemically-laryngealized vowels in languages of the world. Yet creaky voice may well co-occur with glottal stops. For example, phrase-final creak occurs in English (Kreiman, 1982; Redi & Shattuck-Hufnagel, 2001; Wolk, Abdelli-Beruh & Slavin, 2012). Glottal stop allophones of /t/, as in the word 'button' [bʌʔn], are also common. If both glottal stops and creaky voice are perceived by drops in F0 and amplitude, will 'button' be confusable with 'bun' in the domain of phrase-final creak? In other words, is 'button' more confusable with 'bun' when it is pronounced as [bʌ̰ʔn̩] compared to [bʌʔn̩]?

Thus, based on the results from previous perception studies, we can hypothesize that (near-)minimal pairs like *button-bun*, where one word in the pair contains a glottal stop, will indeed be more confusable in creak than when no creak is present. However, the opposite hypothesis is also possible, because phrasal creak differs articulatorily from a glottal stop gesture: during creak the vocal folds are largely abducted, and there is generally an increase in airflow (Slifka, 2006). Conversely, by definition a glottal stop involves constriction of the vocal folds and therefore a decrease in airflow. Acoustically, we therefore expect differences between laryngealization and creak in terms of spectral characteristics, and these could aid in distinguishing creak from glottalization. Nonetheless, if listeners do not focus on the spectral characteristics that might distinguish creak from glottal stops, and instead pay attention only to F0 and amplitude, then glottal stop identification could be hindered by the presence of creak.

To test this, in this study English listeners were asked to identify words like 'button' and 'bun' when (1) the former was pronounced with a [t] vs. short/long [ʔ], and (2) when both words were produced with creaky vs. non-creaky phonation.

## 4.2  Method

### 4.2.1  Stimuli

The stimuli consisted of naturally-produced English (near-)minimal word pairs, for which one of the pairs contained a glottal stop allophone of /t/. There were three groups (each containing seven word pairs), and the groups differed in terms of where the glottal stop allophone of /t/ occurs in one of the words in each pair. In the first group of word pairs, which I call the *button-bun* group, one of the words of the pair had a glottal stop before a syllabic nasal (e.g. *button* [bʌ.ʔn̩]), whereas the other had the same vowel, followed by no glottal stop, just a (non-syllabic) nasal, e.g. *bun* [bʌn]. In the second group of word pairs (the *atlas-Alice* group), the word with a glottal stop had [ʔ] in coda position before a sonorant onset [l, n], e.g. *atlas* [æʔ.ləs], whereas the non-glottalized word only differed by the absence of [ʔ], e.g. *Alice* [æ.ləs]. Of course, there are other cues to the contrast in the *atlas-Alice* group. For example, for the word with [ʔ] we expect a lighter [l] and less anticipatory lateral coarticulation on the preceding vowel, whose duration will also be shorter due to the fact that it is in a closed syllable. In both groups, the medial /t/ in the *button-* and *atlas*-type words are almost always realized as [ʔ] in conversational speech. In the third group of word pairs (the *dent-den* group), one of the words in the pair ended with orthographic *–nt*, which is commonly glottalized as [nʔ] or [nʔ͡t̚] (Huffman, 2005; Sumner & Samuel, 2005). (Note that, because [n] and [t] have the same place of articulation, it is unclear what the articulatory difference between [nʔ] and [nʔ͡t̚] is, assuming a consonantal [n] is produced). The other word in the pair differed only in that the coda contained only /n/, with no following /t/. A complete list of the stimuli can be found in Table 4.1.

The stimuli were produced by a phonetically-trained female speaker, who was instructed to say each word in the carrier phrase 'STEVEN was the one who said the word $X$', with focus on the first word. Therefore, no words following the initial word 'Steven' bore a pitch accent. The focus early in the phrase also facilitated creak at the end (thus, on the target word), due to pitch lowering and deaccenting after the focused, nuclear-pitch-accented word. Words with allophonic glottal stops were said once with an unaspirated or lightly aspirated [t] and once with a full glottal stop [ʔ], in order to create two lexical conditions: [t] vs. [ʔ]. Tokens with [t] were included to establish a baseline for the effect of phrase-final creak on word identification. All words were uttered in the carrier with and without phrase-final creak. Because in English it is quite natural to produce phrase-final creak, especially at a low pitch range, the phonetician was instructed to end in slight breathy voice for the non-creaky condition. The glottal stops in the creaky condition were still produced as full [ʔ]. In general, full [ʔ]s were produced with a closure duration of about 40-50 ms. For Groups 1 and 2 (whose allophonic glottal stops occur word-medially), an additional lexical condition was then created by shortening the closure of the glottal stops to 20-25 ms in duration, effectively making the glottal stop ambiguous between full [ʔ] and incomplete [ʔ̞] (which I denote as short [ʔ̆]). Therefore, in the shortened glottal stop condition the glottal stops were about half as long. A summary of the experimental conditions with sample transcriptions is shown in Table 4.2.

The stimuli were recorded at a sampling rate of 22,050 Hz using PCquirer in a sound-attenuated room at UCLA, using a Shure SM10A head-mounted microphone whose signal ran through an XAudioBox pre-amplifier and A-D device. The target words were then extracted from the carrier phrase in Praat (Boersma & Weenink, 2011). Prior to presentation to listeners, all stimuli were equalized for peak amplitude and multiplied by 20-ms ramps to eliminate onset and offset click artifacts. Sample waveforms for the *button* stimuli are shown in Figure 4.1 (p. 60).

| Group 1: 'button-bun' /CV.tn̩/ vs. /CVn/ | Group 2: 'atlas-Alice' /(C)Vt.SV(C)/ vs. (C)V.SV(C)/ | Group 3: 'dent-den' /CVnt/ vs. /CVn/ |
|---|---|---|
| button-bun | atlas-Alice | dent-den |
| rotten-Ron | motley-Molly | jaunt-John |
| gotten-gone | Whitney-whinny | tint-tin |
| beaten-bean | witless-Willis | lint-Lynn |
| Keaton-keen | settler-seller | font-fawn |
| satan-sane | greatness-grayness | paint-pain |
| Dayton-dane | curtly-curly | faint-feign |

Table 4.1: Stimulus word pairs

| Condition | Group 1: 'button-bun' | Group 2: 'atlas-Alice' | Group 3: 'dent-den' |
|---|---|---|---|
| Non-creaky + [t] | [bʌtn̩]-[bʌn] | [ætləs]-[æləs] | [dɛnt]-[dɛn] |
| Non-creaky + [ʔ] | [bʌʔn̩]-[bʌn] | [æʔləs]-[æləs] | [dɛnʔ]-[dɛn] |
| Non-creaky + [ʔ̆] | [bʌʔ̆n̩]-[bʌn] | [æʔ̆ləs]-[æləs] | |
| Creaky + [t] | [bʌ̰tn̩]-[bʌ̰n̰] | [æ̰tlə̰s]-[æ̰lə̰s] | [dɛ̰nt]-[dɛ̰n̰] |
| Creaky + [ʔ] | [bʌ̰ʔn̰̩]-[bʌ̰n̰] | [æ̰ʔlə̰s]-[æ̰lə̰s] | [dɛ̰nʔ]-[dɛ̰n̰] |
| Creaky + [ʔ̆] | [bʌ̰ʔ̆n̰̩]-[bʌ̰n̰] | [æ̰ʔ̆lə̰s]-[æ̰lə̰s] | |

Table 4.2: Summary of experimental conditions with sample IPA transcriptions.
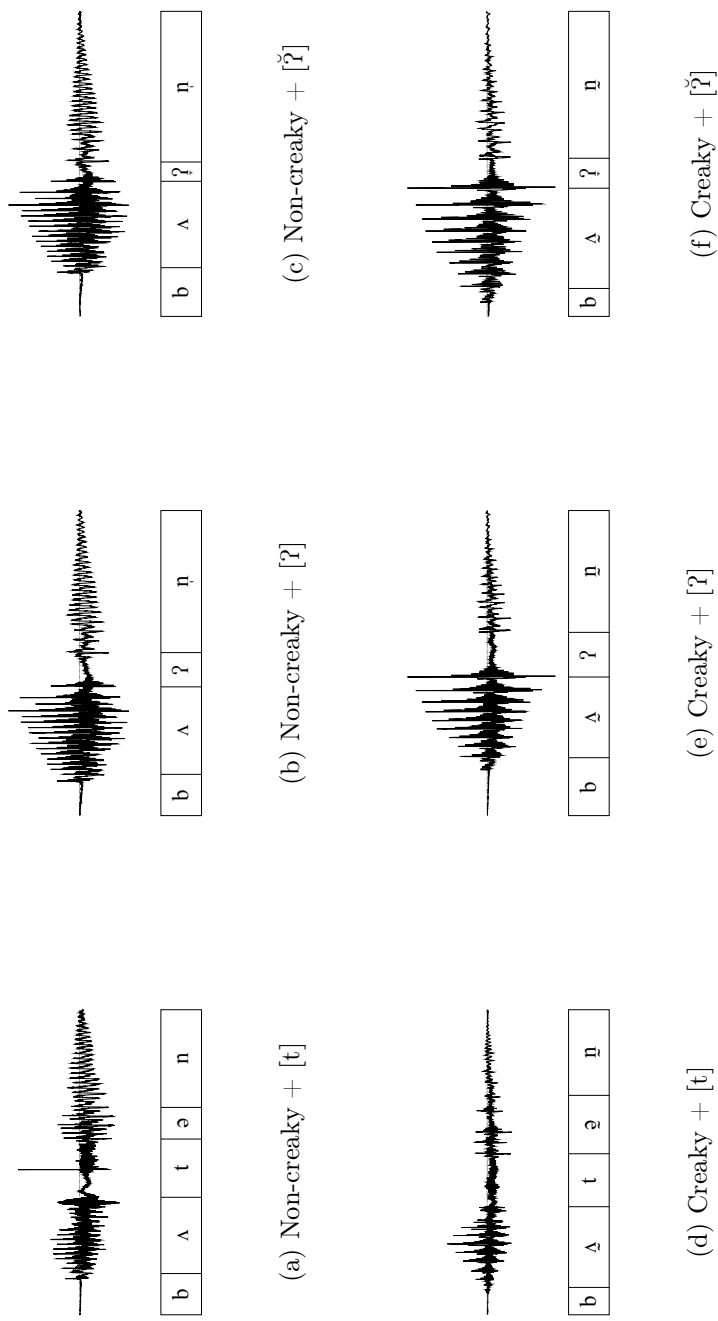
Figure 4.1: Waveforms for the *button* stimuli. The top row are those produced with non-creaky phonation; the bottom row with creaky phonation. The glottal pulses for the creaky stimuli are irregular in both time and amplitude. Note that the stimuli with short glottal stops [ʔ̆] were created by shortening the duration of the glottal closure of the [ʔ] stimuli.

(a) Non-creaky + [t]

(b) Non-creaky + [ʔ]

(c) Non-creaky + [ʔ̆]

(d) Creaky + [t]

(e) Creaky + [ʔ]

(f) Creaky + [ʔ̆]

60

#### 4.2.1.1   Stimuli acoustics

To verify that the creaky and non-creaky stimuli differed acoustically in the manner expected, I analyzed the target stimuli for several common measures of voice quality using VoiceSauce (Shue, Keating, Vicenik & Yu, 2011). Because the non-creaky stimuli were somewhat breathy (which is characterized by higher H1*-H2* values than modal voice (Gordon & Ladefoged, 2001; Garellek & Keating, 2011)), and because phrase-final creak is also expected to have slightly higher H1*-H2* values than modal voice, I expected little difference in spectral slope between the creaky vs. non-creaky stimuli. However, the phonation difference should translate to lower harmonics-to-noise ratio (HNR) for creaky compared to non-creaky stimuli. The breathiness in the non-creaky stimuli could lower HNR due to the higher level of aspiration noise in the signal, but such noise should in theory also be found for voicing during phrase-final creak, which is characterized by increased airflow (Slifka, 2000, 2006). Aspiration noise aside, creaky stimuli should have even lower values of HNR because they are also characterized by signal aperiodicity (increased 'jitter' and 'shimmer'), which is weaker in breathy or modal voice.

The acoustic differences between creaky and non-creaky stimuli were assessed using a logistic mixed-effects regression model, which modeled creaky (vs. non-creaky) stimuli as a function of the average value for 13 measures. Each acoustic measure included as a fixed effect was centered to reduce collinearity, and the model also had a random intercept for word. The results, shown in Table 4.3, indicate that several measures were in fact significant predictors of creakiness in the stimuli. Three spectral tilt measures emerged as significant predictors: H1*-A1*, H1*-A2*, and H1*-A3*. H1*-A1* was higher under creakiness, whereas the other two spectral measures were lower under creakiness. These results can be interpreted physiologically as the result of a more abducted – but more abruptly closing – glottis during phrase-final creak (Hanson et al., 2001).

The other measures to emerge as significant in the model were noise measures: HNR below 500 Hz, below 2500 Hz, below 3500 Hz, and CPP. Interestingly, the HNR measures

|  | Coef $\beta$ | SE($\beta$) | z | p |
|---|---|---|---|---|
| Intercept | 0.31 | 0.54 | 0.6 | >0.6 |
| H1*-H2* | −0.22 | 0.32 | −0.7 | >0.5 |
| H2*-H4* | 0.18 | 0.21 | 0.8 | >0.4 |
| H1*-A1* | 0.56 | 0.23 | 2.4 | **<.05** |
| H1*-A2* | −0.74 | 0.23 | −3.3 | **<.01** |
| H1*-A3* | −0.44 | 0.18 | −2.5 | **<.05** |
| F0 | 0.00 | 0.03 | −0.1 | >0.9 |
| Duration | 0.00 | 0.01 | −0.1 | >0.9 |
| HNR < 500Hz | −0.66 | 0.28 | −2.3 | **<.05** |
| HNR < 1500Hz | −0.02 | 0.46 | −0.1 | >1 |
| HNR < 2500Hz | −2.11 | 0.74 | −2.9 | **<.01** |
| HNR < 3500Hz | 2.73 | 0.73 | 3.7 | **<.001** |
| CPP | 0.98 | 0.48 | 2.0 | **<.05** |
| Energy | −0.20 | 0.16 | −1.3 | >0.2 |

Table 4.3: Results of the logistic model predicting creaky stimuli as a function of 13 acoustic measures.

below 2500 Hz showed a decrease in relative harmonic energy (i.e., lower values) under creakiness, whereas the noise measures calculated over larger frequency ranges (HNR below 3500 Hz and CPP) show a decrease in noise. This may be due to the decrease in periodicity under creak (resulting in more noise in the lower frequencies), but also to the increase in higher-frequency energy caused by more abrupt closure.

These results indicate that there are consistent and expected acoustic differences between the creaky and non-creak conditions. Therefore, if there is indeed an effect of creak on word identification, it is likely to come from its changes to the harmonic and noise components during voicing.

## 4.2.2 Participants

Sixteen monolingual native English listeners (mean age = 20.5, eight women and eight men) were recruited at UCLA. None reported any history of hearing impairment. Listeners received course credit for their participation.

### 4.2.3 Task

Listeners participated in a two-alternative forced-choice task, in which they were instructed to listen to a word over headphones and identify which word they heard. The experiment was implemented in Praat (Boersma & Weenink, 2011). Listeners were presented visually with the target word and its alternative from the word pairs (shown in Table 4.1). They heard the target word only once, after which they selected which word they thought they heard. They then rated their confidence in their choice on a Likert scale from 1 'total guess' to 5 'totally sure'. The listeners could change their word choice and confidence score as many times as they liked before moving on to the next trial.

The first two groups made up 168 experimental tokens: 2 stimulus groups (*button-bun* and *atlas-Alice*) x 7 pairs x 2 words per pair x 2 phonation types (non-creaky and creaky) x 3 /t/ allophones. Words with no medial /t/ (e.g. *bun*) were presented three times so that each item in a word pair would yield the same number of tokens. The third group (*dent-den*) made up 56 experimental tokens: 7 pairs x 2 words per pair x 2 phonation types (non-creaky and creaky) x 2 /t/ allophones. In total, 224 trials were therefore included, and each trial was repeated twice in separate blocks. The words were presented randomly within block. The experiment lasted about 20-30 minutes.

## 4.3 Results

### 4.3.1 Effect of vowel quality

Several pairs of stimuli (e.g., *fawn* /fɔn/ vs. *font* /fɑnt/) differ in vowel quality in some varieties of English, though they were produced by a speaker with the so-called 'caught-cot' (/ɔ–ɑ/) merger. Even though all participants were native speakers of Californian English, and thus are supposed to have the merger, a listener who (for whatever reason) does not might assume that the token heard would be the word with /ɑ/, since the speaker never produced instances of [ɔ]. To ensure that a vowel mismatch (in any variety of English) did

not affect accuracy, I ran a logistic mixed-effects model predicting accuracy as a function of the presence vs. absence of a potential vowel mismatch between stimulus pairs (with random intercepts for participants and items, and random slopes for participants and items for potential vowel mismatch). The results reveal no significant effect of potential vowel mismatch on accuracy ($\beta = 1.42$, $z = 1.14$, $p = 0.25$). Thus, the subsequent analysis does not distinguish between stimulus pairs with a potential vowel mismatch and those without.

### 4.3.2 Words with glottal allophones: *button, atlas*, and *dent*

Next we assess the effect of /t/ allophone and phonation type on accuracy and confidence. The results are organized according to stimulus group. For the *button* and *atlas* words, subsequent pairwise (non-creaky vs. creaky) comparisons for each segment type ([t], [ʔ], and [ʔ̆]) were run using backward difference contrast coding, such that the mean of each bar in a figure is compared to the mean of the previous bar.[1] These models had random intercepts for listener and stimulus. For the linear mixed-effects models (for confidence), the *p*-values are obtained using MCMC sampling with 10,000 iterations (Baayen, 2008), whereas the *p*-values for the logistic models are provided in the model summary. Because accuracy was at ceiling (with zero variance) for the [t] responses (cf. Figures 4.2a, 4.3a, and 4.4a), one data point (with an inaccurate response) was added for each phonation-by-allophone permutation (for the three stimulus pair types), so that the logistic regression model would not have inflated standard errors. However, the results shown in Figures 4.2a, 4.3a, and 4.4a represent only the actual obtained data.

### 4.3.2.1 *Button*-type words: medial glottal stop followed by syllabic nasal

I expect listeners to be good at identifying *button*-type words when they have glottal stops (i.e., good at not mistaking them for *bun*-type words), because they differ from *bun*-type

---

[1] For more details on backwards difference coding, see `http://www.ats.ucla.edu/stat/r/library/contrast_coding.htm#backward`. The models for *dent*-type words and for the words with no glottal allophone will be discussed in Sections 4.3.2.3 and 4.3.3, respectively.

words in terms of the number of syllables, the syllabicity of the final nasal, and because there are available cues into and out of the glottal stop.

Accuracy (proportion correct) and confidence for these words are shown in Figure 4.2. For the tokens with [t] allophone, the accuracy is at ceiling for both non-creaky and creaky tokens, with zero variance. Thus, presence of creak does not hinder identification of sounds in general. For the tokens with either a short or long glottal stop allophone, accuracy is still at or just below ceiling. The within-subject results are shown in Figure B.1 in Appendix B. Nine of 16 participants were at ceiling regardless of the /t/ allophone and phonation, suggesting that these word pairs are easily discriminable, even under creak. A summary of the statistical models for both accuracy and identification confidence is shown in Table 4.4. When the *button*-type words have a (long) glottal stop, there is a slight decrease in accuracy for creaky compared to non-creaky tokens, but this difference is not significant ($p = 0.61$). The same is true when the glottal stop closure duration is shortened, though the magnitude of the difference between non-creaky and creaky tokens increases slightly ($p = 0.09$). In terms of confidence scores, when the glottal stop is shortened and under creak, listeners report lower confidence in their decision compared with their confidence for the non-creaky shortened glottal stops ($p < 0.05$).

### 4.3.2.2 *Atlas*-type words: medial glottal stop followed by sonorant-initial syllable

For the *atlas*-type words, recall that we expect a more detrimental effect of creak on word identification than for the *button*-type words, because the *atlas-Alice* pairs do not differ in syllable count. Thus, there are fewer cues to the *atlas-Alice* contrast compared with the *button-bun* one. Accuracy (proportion correct) and confidence for these words are shown in Figure 4.3. (Within-subject accuracy results are shown in Figure B.2 in Appendix B).

|  | Coef $\beta$ | SE($\beta$) | z | p |
|---|---|---|---|---|
| Intercept | 4.57 | 0.44 | **10.21** | **<.0001** |
| Creaky vs. non-creaky [t] | −0.37 | 1.80 | −0.21 | 0.84 |
| Creaky vs. non-creaky [ʔ] | −0.77 | 1.49 | −0.51 | 0.61 |
| Creaky vs. non-creaky [ʔ̆] | −1.85 | 1.08 | −1.72 | 0.09 |

(a) Accuracy

|  | Coef $\beta$ | SE($\beta$) | t | p (MCMC) |
|---|---|---|---|---|
| Intercept | 4.80 | 0.07 | **64.25** | **<.0001** |
| Creaky vs. non-creaky [t] | 0.00 | 0.12 | 0.00 | 0.99 |
| Creaky vs. non-creaky [ʔ] | −0.06 | 0.12 | −0.53 | 0.56 |
| Creaky vs. non-creaky [ʔ̆] | −0.14 | 0.12 | **-2.33** | **<.05** |

(b) Confidence

Table 4.4: Relevant pairwise comparisons for accuracy and confidence of identification of *button*-type words.

As was found for the *button*-type words, accuracy for the *atlas*-type words is at ceiling for both non-creaky and creaky tokens with [t], with zero variance, indicating that the presence of creak does not hinder identification of sounds in general. A summary of the statistical models for both accuracy and identification confidence are shown in Table 4.5. When the *atlas*-type words have a (long) glottal stop, there is a significant decrease in accuracy for creaky compared to non-creaky tokens ($p < 0.05$). The same is true when the glottal stop closure duration is shortened ($p < 0.01$). Indeed, the within-subject plots (Figure B.2) show that half the participants were around or below chance when identifying *atlas*-type words with shortened glottal stops.

The confidence scores mirror the accuracy results: for both [ʔ] and [ʔ̆] segment types, confidence is lower for creaky than for the non-creaky counterparts ($p < 0.001$ for [ʔ] and $< 0.0001$ for [ʔ̆]). Thus, we see that the *atlas*-type words are harder to identify under creak (relative to their *Alice*-type counterparts) than the *button*-type words. For the latter, only when the glottal stop is shortened is there any significant detriment, and only in terms of rater confidence. For the *atlas*-type words, however, there is a significant decrease in accuracy and confidence when the medial /t/ is [ʔ] or shortened [ʔ̆].

66
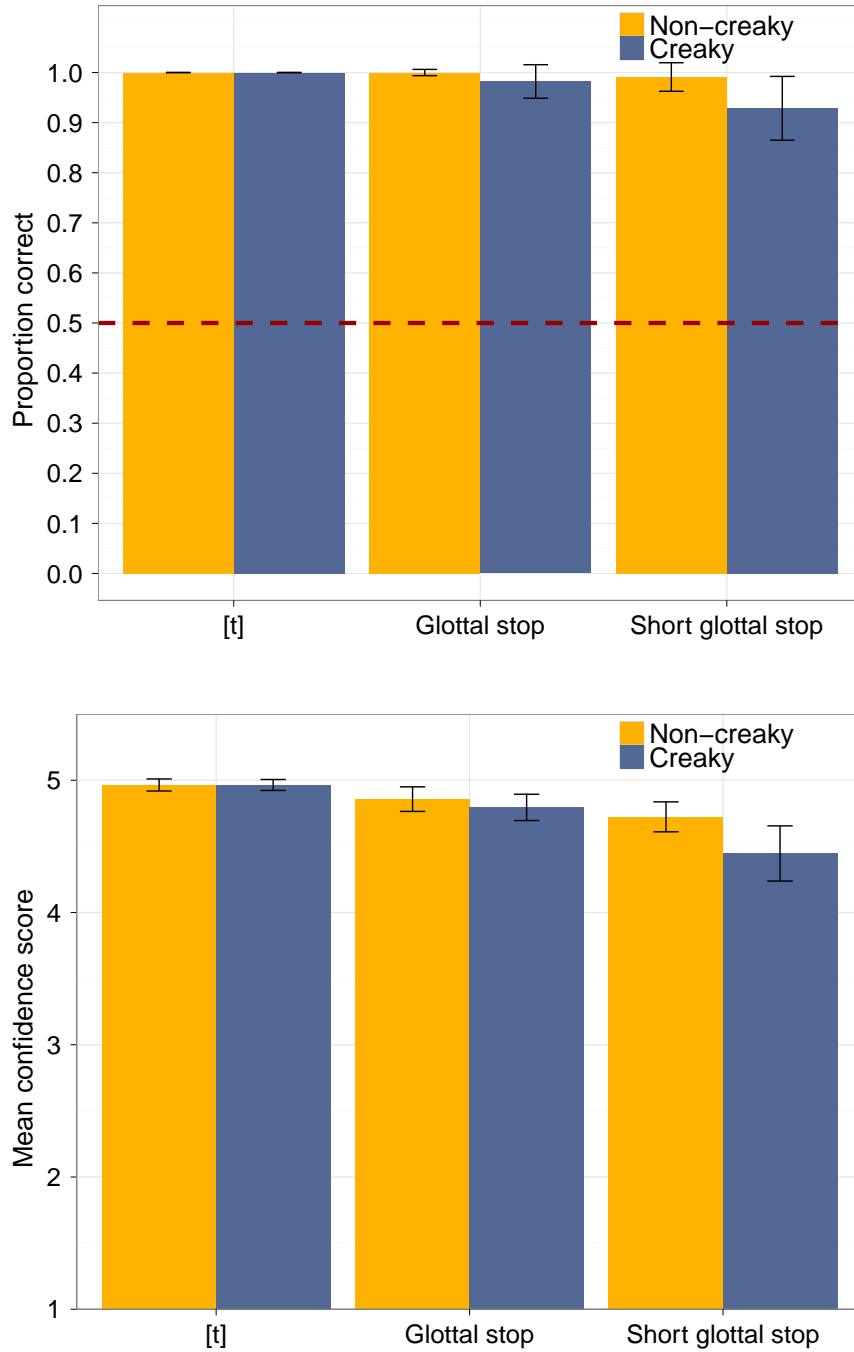
Figure 4.2: Mean proportion correct (top) and confidence (bottom) for *button*-type words. For proportion correct, chance (50%) is marked by a dashed line. A higher confidence rating signifies increased confidence in the participant's choice. Error bars indicate 95% confidence intervals.

|  | Coef $\beta$ | SE($\beta$) | z | p |
| --- | --- | --- | --- | --- |
| Intercept | 3.86 | 0.49 | **7.82** | **<.0001** |
| Creaky vs. non-creaky [t] | −0.02 | 1.98 | −0.51 | 0.61 |
| Creaky vs. non-creaky [ʔ] | −2.13 | 1.05 | **-2.02** | **<.05** |
| Creaky vs. non-creaky [ʔ̆] | −2.93 | 0.93 | **-3.15** | **<.01** |

(a) Accuracy

|  | Coef $\beta$ | SE($\beta$) | t | p (MCMC) |
| --- | --- | --- | --- | --- |
| Intercept | 4.49 | 0.08 | **59.75** | **<.0001** |
| Creaky vs. non-creaky [t] | −0.06 | 0.12 | −0.50 | 0.59 |
| Creaky vs. non-creaky [ʔ] | −0.52 | 0.12 | **-4.13** | **<.001** |
| Creaky vs. non-creaky [ʔ̆] | −0.54 | 0.12 | **-4.27** | **<.0001** |

(b) Confidence

Table 4.5: Pairwise comparisons for accuracy and confidence of identification of *atlas*-type words.
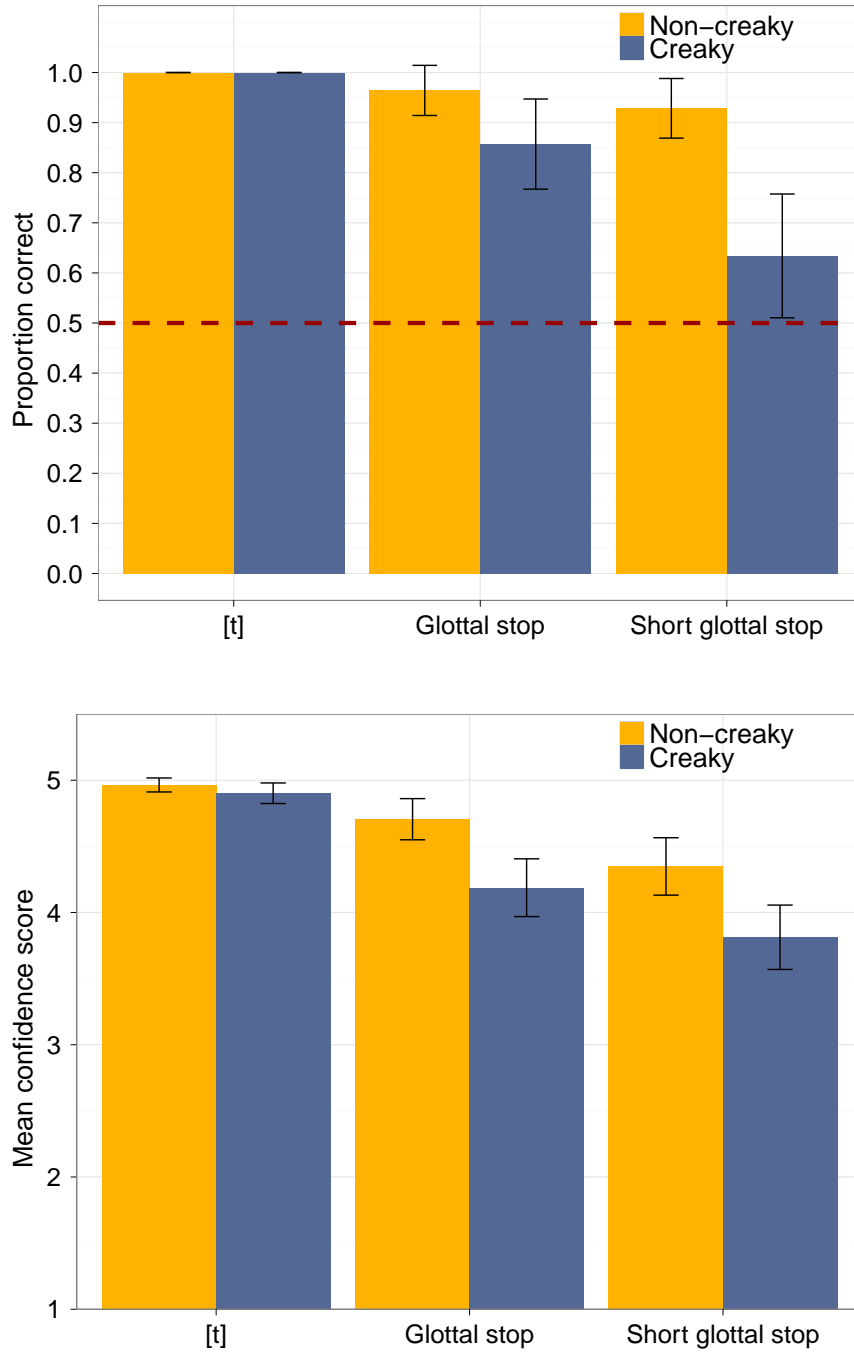
Figure 4.3: Mean proportion correct (top) and confidence (bottom) for *atlas*-type words. For proportion correct, chance (50%) is marked by a dashed line. A higher confidence rating signifies increased confidence in the participant's choice. Error bars indicate 95% confidence intervals.

### 4.3.2.3   *Dent*-type words: word-final glottal stop

For *dent*-type words, I expect that presence of a glottal stop will result in more cases of misidentification (regardless of the presence of creak), because its only (lead) cues are during a nasal. However, the effect of creak should not be additionally harmful if glottal stops and creak may be heard by listeners as instances of glottal stops. Accuracy (proportion correct) and confidence for these words are shown in Figure 4.4. (Accuracy results within subject are shown in Figure B.3 in Appendix B). Recall that, because the glottal stops were always word-final, there was no shortened [ʔ̆] condition.

The effects of phonation (no creak vs. creak) and segment type ([t] vs. [ʔ]) were assessed by mixed-effects regression predicting accuracy/confidence as a function of phonation, segment type, and their interaction. The models had random intercepts for listener and stimulus and random slopes for phonation. A summary of the statistical models for both accuracy and identification confidence is shown in Table 4.6. The accuracy results indicate that the presence of a glottal stop results in a significant decrease in accuracy relative to the words with [t] ($p < 0.01$), but there was no significant effect of phonation and no significant interaction between phonation and segment type. Indeed, the within-subject plots (Figure B.3) reveal that three participants were at or below chance for the glottal stop condition, regardless of the phonation condition. On the other hand, five participants were at ceiling in all conditions, meaning that some listeners are good at detecting post-nasal [ʔ], whether the nasal is produced with creak or not.

For identification confidence, significance is assumed if the absolute $t$-value is greater than 2, because $p$-values cannot be calculated for linear models with random slopes (Baayen, 2008). There is a significant interaction between phonation and segment type, with creaky-[ʔ] tokens having the lowest identification confidence overall. Tokens with [ʔ] are identified with less confidence than those with [t], and the main effect of phonation is also significant, with creaky tokens having lower identification confidence than non-creaky tokens.

|  | Coef $\beta$ | SE($\beta$) | z | p |
|---|---|---|---|---|
| Intercept | 6.53 | 1.29 | **5.07** | **<.0001** |
| Phonation = creaky | −0.24 | 1.85 | −0.13 | 0.90 |
| Allophone = [ʔ] | −3.784 | 1.24 | **-3.05** | **<.01** |
| Phonation = creaky*Allophone = [ʔ] | −0.35 | 1.97 | −0.18 | 0.86 |

(a) Accuracy.

|  | Coef $\beta$ | SE($\beta$) | t |
|---|---|---|---|
| Intercept | 3.85 | 0.15 | **25.54** |
| Phonation = creaky | −0.30 | 0.14 | **-2.15** |
| Allophone = [ʔ] | 0.97 | 0.16 | **6.07** |
| Phonation = creaky*Allophone = [ʔ] | 0.38 | 0.18 | **2.08** |

(b) Confidence.

Table 4.6: Effects of phonation, allophone type, and their interaction on accuracy and confidence of identification of *dent*-type words.

### 4.3.3 *Bun-, Alice-,* and *den*-type words: no allophonic [ʔ]

Thus far, it is evident that English listeners' identification of glottal stops is hindered by the presence of phrase-final creak. This implies either that creak is misheard as a glottal stop, or that glottal stops are harder to identify in creak, or both. Because listeners were also presented with words with no /t/, so no allophonic glottal stops (i.e., the *bun, Alice,* and *den* words from the *button-bun, atlas-Alice*, and *dent-den* pairs), we can therefore determine if creak is sometimes misheard as a glottal stop.

The overall results for accuracy (proportion correct) and identification confidence are shown in Figure 4.5, and within-subject plots for accuracy can be found in Figure B.4 in Appendix B. For these words, subsequent pairwise (non-creaky vs. creaky) comparisons for each stimulus type (*bun, Alice,* and *den*) were run using backward difference contrast coding, such that the mean of each bar in a figure is compared to the mean of the previous bar. These models had random intercepts for listener and stimulus. For the linear mixed-effects models (for confidence), the *p*-values are obtained using MCMC sampling with 10,000 iterations (Baayen, 2008). The *p*-values for the logistic models (for accuracy) are provided in the model
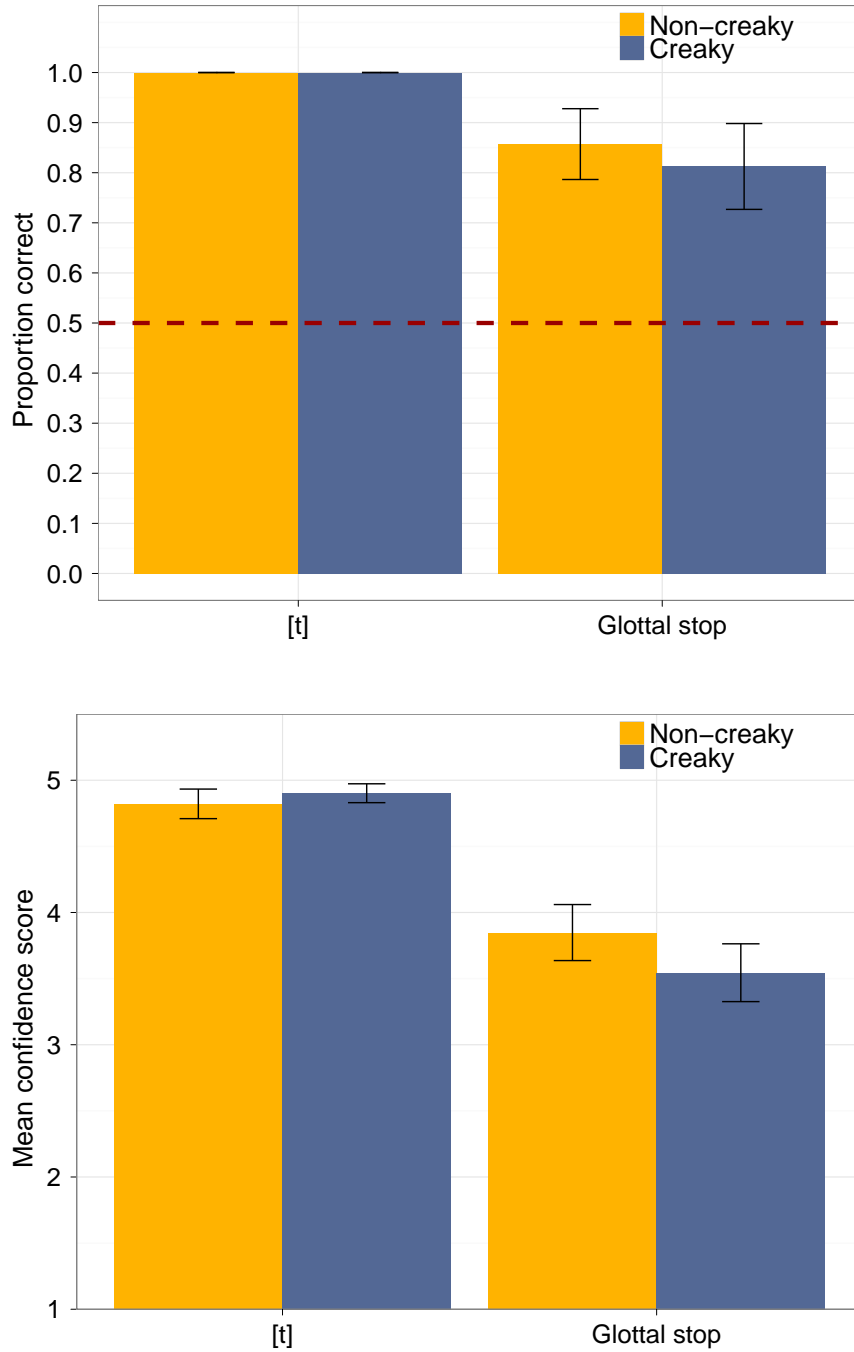
Figure 4.4: Mean proportion correct (top) and confidence (bottom) for *dent*-type words. For proportion correct, chance (50%) is marked by a dashed line. A higher confidence rating signifies increased confidence in the participant's choice. Error bars indicate 95% confidence intervals.

summary. The results (see summaries in Table 4.7) indicate that the presence of creak does not result in a decrease in accuracy for *bun-* and *Alice*-type words. This pattern is fairly robust across participants, with eleven of the 16 participants at or near ($> 90\%$) ceiling (see Figure B.4). However, *den*-type words were less accurately identified when creaky ($p < 0.01$), with two participants at or below chance (see Figure B.4). That is, listeners think there is a /t/, presumably as [ʔ]. Thus, we can conclude that when other cues to the distinction between glottal stops and creak are weak (e.g., only during a nasal), creak can be mistaken for a glottal stop. Generally, though, identification confidence is lower for words with creak ($p < 0.0001$ for the three stimulus types), though they still remain on the higher ('totally sure') side.

|  | Coef $\beta$ | SE($\beta$) | z | p |
|---|---|---|---|---|
| Intercept | 5.08 | 0.46 | **11.12** | **<.0001** |
| Creaky vs. non-creaky *bun* | $-1.86$ | 1.67 | $-1.10$ | 0.27 |
| Creaky vs. non-creaky *Alice* | $-1.39$ | 1.26 | $-1.10$ | 0.27 |
| Creaky vs. non-creaky *den* | $-4.82$ | 1.60 | **-3.01** | **<.01** |

(a) Accuracy

|  | Coef $\beta$ | SE($\beta$) | t | p (MCMC) |
|---|---|---|---|---|
| Intercept | 4.63 | 0.06 | **72.83** | **<.0001** |
| Creaky vs. non-creaky *bun* | $-0.36$ | 0.08 | **-4.36** | **<.0001** |
| Creaky vs. non-creaky *Alice* | $-0.41$ | 0.08 | **-4.88** | **<.0001** |
| Creaky vs. non-creaky *den* | $-0.64$ | 0.09 | **-7.28** | **<.0001** |

(b) Confidence

Table 4.7: Main effects of phonation and stimulus type for accuracy and confidence of identification of *bun-*, *Alice-*, and *den*-type words.
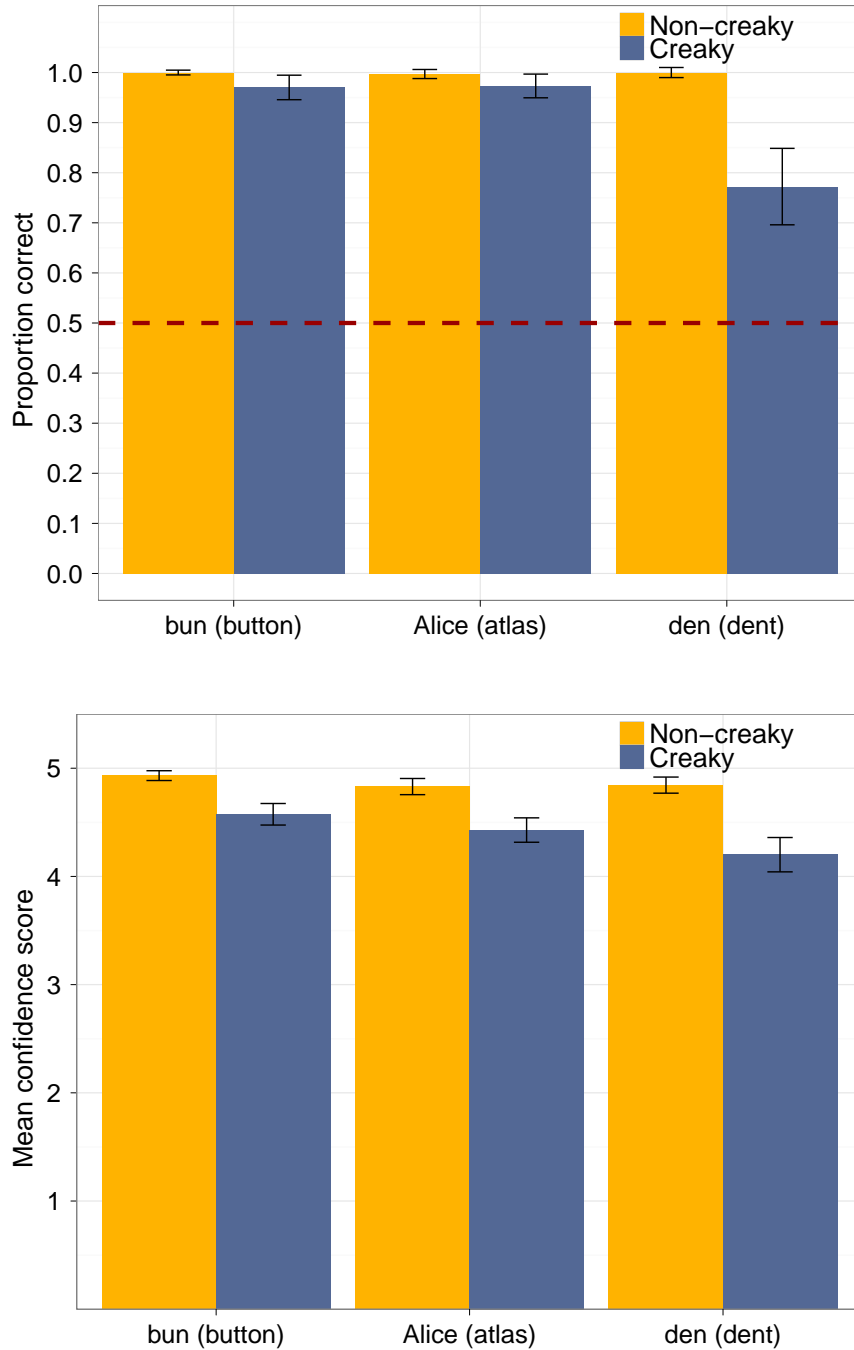
Figure 4.5: Mean proportion correct (top) and confidence (bottom) for words with no /t/ and thus no allophonic glottal stop. The incorrect percepts are noted in parentheses. For proportion correct, chance (50%) is marked by a dashed line. A higher confidence rating signifies increased confidence in the participant's choice. Error bars indicate 95% confidence intervals.

## 4.4 Discussion

### 4.4.1 Summary of results

The main goal of this study was to determine whether glottal stop identification is hindered by the presence of creak. This is partially confirmed by the results: listeners are generally poorer at identifying words with glottal stops when they occur in creak than when no creak is present. However, overall accuracy and confidence remain high, and the influence of creak on word identification varies depending on the length of the glottal stop and the type of word. The effects of glottal allophone and word set on accuracy and confidence are summarized in Table 4.8.

| | Accuracy | Confidence |
|---|---|---|
| *button-bun* pairs | | Creaky *button*: ↓ when [ʔ̆] <br> Creaky *bun*: ↓ |
| *atlas-Alice* pairs | Creaky *atlas*: ↓ when [ʔ, ʔ̆] | Creaky *atlas*: ↓ when [ʔ, ʔ̆] <br> Creaky *Alice*: ↓ |
| *dent-den* pairs | *dent*: ↓ when [ʔ] <br> Creaky *den*: ↓ | *dent*: ↓ when [ʔ] and/or creaky <br> Creaky *den*: ↓ |

Table 4.8: Summary of results.

In this study, the adverse effect of creak is particular to [ʔ] identification; creak does not influence word identification when the /t/ is realized as [t] instead of a glottal stop. Thus, creak is not intrinsically detrimental to word identification. Rather, it is detrimental for glottal stops in particular, most likely because the acoustic properties of creak resemble those of glottal stops.

75

The target words with glottal allophones (*button-, atlas-*, and *dent*-type words) are in general more accurately identified depending on the number of possible cues available in the signal. Thus, identification of *button*-type words is at ceiling or just below, even when the /t/ allophone is a shortened glottal stop (see Figure 4.2a). Presumably, this is because there are several potential cues allowing listeners to differentiate such words from their alternatives (e.g., *bun*): changes into and out of the glottal stop, as well as syllabicity. On the other hand, identification of *atlas*-type words is just above chance when the /t/ allophone is a shortened glottal stop (see Figure 4.3a). Such words form minimal pairs with their paired words, such that the only difference between *atlas* and *Alice* is the presence of a /t/ in the former but not in the latter. Therefore, the absence of a difference in syllabicity between the paired words is perceptually costly, but only in creak, and only when the realization of the medial /t/ is a (short) glottal stop.

The *dent*-type words were expected to be the most affected by creak when glottalized, because unlike *button-* and *atlas*-type words, they lack segments following the glottal stop. This hypothesis was not confirmed; identification accuracy for the *dent*-type words when glottalized (e.g., [denʔ]) mimicked that of the glottalized *atlas*-type words (e.g., [æʔləs]). It is also worth noting that the *dent*-type words were the only group for which [ʔ] caused a drop in accuracy when *no creak* was present (cf. accuracy for non-creaky [t] vs. non-creaky [ʔ] in Figure 4.4a). Thus, glottalization affects accuracy for *dent*-type words *regardless* of creak, but for *button-* and *atlas*-type words, identification is at ceiling for the glottal stop variant, or just below ceiling for the short glottal stop variant. That glottalization should be detrimental regardless of creak is not surprising, because the only potential differences between *dent*-type words and their *den*-type counterparts are shorter rhymes and irregular voicing at the end (which may be wrongly attributed to phrase-final creak). The irregular voicing occurs during the nasal coda, and nasals are generally weaker sounds than vowels. Thus, irregular voicing during a nasal is probably less perceptible than during a vowel. However, it is also possible that the distinction between creak and glottalization is harder to

hear word-finally (regardless of what segment type they appear on), because word-final [ʔ] only occurred after nasals in the stimuli.

### 4.4.2 Is creak misheard as glottalization?

Are words with no glottal stop sometimes misidentified in creak? For example, do listeners confuse creaky 'bun' [b̰ʌ̰n̰] for *button*? This is important because it enables us to determine if glottal stops and creak are mutually confusable or if glottal stops are simply harder to identify in the presence of creak.

For words like *bun* and *Alice*, which contrast with words like *button* and *atlas*, respectively, the presence of creak does little to hinder word identification (see Figure 4.5a), though there is a slight decrease in confidence as a function of creak (Figure 4.5b). On the other hand, *den*-type words do show a significant decrease in accuracy under creak. This finding suggests that when creaky-voiced nasals occur word-finally (e.g., *den* pronounced as [dɛn̰]), they can sometimes be misheard as belonging to an underlying /nt#/ cluster being realized as [nʔ#]. However, words like *bun* or *Alice* in phrase-final creak are not generally misheard as *button* or *atlas*. Thus, whereas glottal stops (especially when short) are regularly misheard as creak, only creaky word-final nasals are sometimes misheard as underlying /nt/.

### 4.4.3 On the perception of glottal stops vs. creak

The results of this study reveal that most listeners do not typically mistake glottal stops for creak, or creak for glottal stops, when there are ample cues that can be used to distinguish the two gestures (e.g., in words like *button* and *bun*). Errors in word identification depend largely on the length of the glottal stop (in words with glottal stops), and the number of available cues to the glottal stop-creak distinction. Therefore, the results imply that, when perceiving words in creak and/or words with glottal stops, listeners do not rely solely on dips in F0 and amplitude. Rather, they likely use other cues in differentiating creaky words from words with glottal stops. What might these cues be? In this study we are limited by the small

number of tokens used for stimuli, but paired t-tests reveal that creaky segments (vowels or nasals) before glottal stops (e.g., **bu*tt*on, de*nt***) had higher energy and noise (based on HNR measures and CPP) than creaky vowels or nasals not followed by a glottal stop (e.g., *bun, den*). Creaky segments not followed by [ʔ] were also longer in duration than those followed by [ʔ] (significant t-test results are given in Table C.1 in Appendix C).

Thus, most listeners are able to recover glottalization in creak, and to identify creak instead of glottalization, likely via the cues that were found to distinguish glottal stops in creak from plain creak. However, they are better at recovering glottalization when (1) no creak is present, (2) when cues to glottalization can be robustly realized (vs. when glottalization is short, or if it occurs after nasals).

The largest detriment to glottal stop recoverability was for short glottal stops when they occurred in *atlas*-type words (see Figure 4.3a). When these forms occurred in creak, the performance of half the subjects was around or below chance (Figure B.2). Thus, it appears that some listeners pay closer attention to cues that could be used to disambiguate creak from glottalization than do others. This is not surprising, because the spectral cues to glottalization (e.g., lower spectral tilt in neighboring voicing) are not contrastive in English. However, it is possible that listeners of a language with contrastive laryngealization would be better at retrieving cases of short glottal stops in creak.

## 4.5 Conclusion

Previous studies have shown that listeners may rely on similar cues when perceiving glottalization and creaky voice. The goal of the present study was therefore to determine whether words with glottal stops would be hard to identify when they occur in the domain of phrase-final creak. The results from an English word identification task reveal that, in creak, words with glottal stops are not misidentified, *except* when few cues are available to disambiguate these words from minimal pairs with no glottalization, and when the glottal stop is short.

Nonetheless, listeners are consistently less sure of their decisions when creak is present, even when word identification is accurate.

# CHAPTER 5

# Word-initial glottalization:
# Where and why does it occur?

## 5.1 Introduction

### 5.1.1 Optional word-initial glottalization: a near-universal property of language?

Glottal stops occur before vowel-initial words in many languages (e.g., English 'apple' pronounced as [ʔæpl]). In some (e.g., Arabic, Ilokano), glottal stops are thought to obligatorily mark word-initial vowels, whereas in others, like English, the presence of word-initial glottal stops is more variable (Lombardi, 2002; Borroff, 2007; Hayes, 2009). This variable phenomenon, which I refer to here as 'word-initial glottalization,' might occur in all languages except those that contrast /#ʔV/ and /#V/. For example, word-initial glottalization is banned in Tongan, where words like /aa/ 'heat sticks over fire' and /ʔaa/ 'awake' contrast. It is indeed remarkable that word-initial glottalization is so common a phenomenon: no other segment is used epenthetically in the same environment and in so many languages. The goal of this chapter is to understand why glottal stops are typologically common before vowel-initial words.

A theory of word-initial glottalization should ultimately account for the phenomenon's cross-language prevalence. But for such a theory to be valid, its underlying assumptions need validation. For example, in my earlier statement, ('no other segment is used epenthetically in the same environment and in so many languages'), I assumed that glottal stops are segments

(equivalent to, e.g, [t]), and that they are inserted, rather than arising through fortition of preexisting segments or features, a distinction that I will explain later. Further, I assumed that their reason for occurring before vowel-initial words is the same across languages. These assumptions are proposed, if not adopted, in many studies of glottalization (Pierrehumbert & Talkin, 1992; Dilley et al., 1996; Borroff, 2007), but are they justified? For example, if we treat glottal stops as segments, we must ask why researchers find that they are usually (even in prosodically-strong environments) realized non-canonically, as a form of irregular voicing or laryngealization (Pierrehumbert & Talkin, 1992; Dilley et al., 1996). Why call them glottal 'stops' if they are rarely plosives? For this reason, the term 'glottalization' is often used to refer to both voicing irregularity (irregular pitch periods) and glottal stops (Redi & Shattuck-Hufnagel, 2001). This in turn leads to another assumption, namely that all instances of glottalization can be treated as reflexes of a single phenomenon.

Even if not all instances of voicing irregularity are derived from the same phenomenon, it is clear that word-initial glottalization at least sometimes results in an incomplete glottal stop ([ʔ̞], a form of laryngealization), and this too has implications for theories of word-initial glottalization. If glottal stops are epenthetic, then we must reconcile the fact that the speaker goes to the added effort to epenthesize, while failing to realize this inserted segment in its canonical form.

In sum, previous attempts to characterize word-initial glottalization have made important theoretical assumptions that have yet to be validated: is word-initial glottalization to be treated as epenthesis? If so, of a segment or a feature? What should we count as a token of word-initial glottalization? In order to develop a valid theory of word-initial glottalization, I believe we should first determine what should count as an instance of glottalization, which in turn will allow us to determine the most important factors in predicting when the phenomenon occurs. Finally, we can then determine why word-initial glottalization occurs, and if it can be treated equally across languages, and if so, why it is so typologically common.

### 5.1.2 When does word-initial glottalization occur?

Although cross-linguistically widespread, it is clear that the frequency of occurrence of word-initial glottalization may differ across languages. For example, it is thought to be rare in Spanish, common in English and German, and almost across-the-board in Czech (Bissiri & Volín, 2010; Bissiri, Lecumberri, Cooke & Volín, 2011; Pompino-Marschall & Żygis, 2011). Many researchers have investigated (for a variety of languages) the factors that promote the occurrence of word-initial glottalization. (Here I discuss only word-initial glottalization. For factors that affect coda glottalization or glottal replacement, see Milroy, Milroy, Hartley & Walshaw (1994); Pierrehumbert (1995); Huffman (2005); Eddington & Channer (2010) and Eddington & Savage (2012), among others.) Predictors of word-initial glottalization may be segmental, lexical, prosodic, or sociolinguistic. In English, segmental factors include hiatus (V#V) environments (Umeda, 1978; Dilley et al., 1996; Pierrehumbert, 1995; Mompeán & Gómez, 2011; Davidson & Erker, 2012) and word-initial back vowels are found to glottalize more frequently than non-back vowels (Umeda, 1978). As for lexical factors, content words exhibit more frequent glottalization than function words (Umeda, 1978). Women are known to use glottalization more than men (Byrd, 1994; Dilley et al., 1996). Most studies emphasize the role of prosody, with presence of stress and/or a pitch accent on the word-initial vowel or later in the word, as well as a larger disjuncture with the preceding word, as factors favoring the occurrence of glottalization (Pierrehumbert & Talkin, 1992; Pierrehumbert, 1995; Dilley et al., 1996). In other languages, additional factors that promote the occurrence of word-initial glottalization include word length in Dutch (Jongenburger & van Heuven, 1991), presence of a preceding pause (Kohler, 1994) as well as speech rate (all of which are correlated with other prosodic factors), and low vowel quality (Brunner & Żygis, 2011; Pompino-Marschall & Żygis, 2011) for German.

Despite abundant interest in the topic, there has been little investigation of the relative importance of the various factors that promote word-initial glottalization. That is, we know what factors play a role in the occurrence of word-initial glottalization, but we do not know how important they are compared to each other. Again, some factors are correlated with

others, e.g. vowel hiatus and pitch accent (Redi & Shattuck-Hufnagel, 2001), changes in speech rate are correlated with changes in prosody, and the distinction between function vs. content words correlates with differences in lexical frequency. A model predicting the occurrence of word-initial glottalization should take into account all these factors, and determine which are most important, so that we can understand why it occurs.

### 5.1.3 Why is word-initial glottalization so common?

If, as mentioned in the preceding section, prosody plays an important role in determining when word-initial glottalization occurs, it is likely that the phenomenon is (sometimes) a form of prosodic strengthening of word-initial vowels (Fougeron, 2001). Prosodic strengthening is the process by which articulations are 'strengthened' in prosodically strong environments, notably phrase-initially and under prominence (Keating, 2006). By 'strengthening,' the articulation itself can become more forceful (Fougeron, 2001), or the contrast between the target and neighboring sounds can be enhanced (Hsu & Jun, 1998; Cho & Jun, 2000; Cho, 2005). The former is usually called 'paradigmatic enhancement,' in contrast to the latter, which is 'syntagmatic.'

If word-initial glottalization is a form of prosodic strengthening, it is unclear what in fact is being strengthened. Is voice quality generally more forceful in strong environments? More forcefully-articulated voicing should yield laryngealization, which has increased glottal closure (Gordon & Ladefoged, 2001), and there is some evidence that earlier portions of an English utterance has tenser voice quality compared to later portions (Epstein, 2002). Strengthening of voice quality provides a straightforward account of why word-initial glottalization is cross-linguistically common, because a more forceful articulation of voicing should result in laryngealization in every language. But if voicing in general is strengthened, then two important facts should be considered. First, in terms of theory, glottal stops should then not be regarded as distinct segments, but as the extreme result of voicing strengthening, as discussed by Pierrehumbert & Talkin (1992); Dilley et al. (1996), and Borroff (2007), among others. Thus, an incomplete glottal stop should not be regarded as a form of lenition, be-

cause it is stronger than the default – modal voicing. Rather, a full [ʔ] is the extreme case of this fortition (see discussion of similar issues regarding English /t/ affrication by Buizza & Plug (2012)). Second, in terms of theory prediction (though never discussed in previous studies), if word-initial glottalization is a form of voice quality strengthening, then all voiced sounds, including voiced consonants, should be expected to show increased laryngealization or even glottal stops in strong environments. However, if vowels are strengthened differently than other voiced sounds, then word-initial glottalization (by means of glottal stop insertion) would be viewed as specific to word-initial vowels. Under this view, a glottal stop might be obligatorily or optionally inserted before all vowel-initial words, and lenited to modal voicing in the weakest prosodic environments. Note that it might be impossible to determine phonetically whether word-initial glottalization is obligatory or optional, if we posit that lenited glottal stops can be realized as modal voicing in the weakest environments. Nonetheless, if glottal stops are always present (but lenited in weak environments), then this has problematic implications for typology and phonological analyses: it would then be unclear why word-initial glottalization is typologically common, and why 'optional' word-initial glottal stops do not pattern with obligatory glottal stops (Borroff, 2007).

In sum, there are important, hitherto unanswered questions regarding glottal stops: are they caused by extreme strengthening of voicing, or are they inserted stops (with incomplete variants in less extreme cases)? And how can we disambiguate these two scenarios?

### 5.1.4   The current study

The goals of this paper are thus to determine (1) whether word-initial glottalization is largely a prosodic phenomenon, and if so, (2) whether the phenomenon arises from strengthening of voice quality or from glottal stop insertion that is specific to word-initial vowels. I address these goals in two studies. In Section 5.2, I analyze a corpus of American English to determine where full glottal stops ([ʔ]) occur. I focus on full glottal stops because other cases of irregular voicing might be due to other sources (Garellek, 2012). In the corpus study, I also determine whether the same environments show acoustic signs of laryngealization (with increased vocal

fold closure) when no glottal stop is present. In Section 5.3, I use a combination of acoustic and articulatory analyses to confirm the acoustic results from the first study. In Section 5.4, I synthesize the findings from these two studies, and propose a revised prosodic account of word-initial glottalization.

## 5.2 Corpus analysis of word-initial glottal stops

In this study, I investigate which factors best predict word-initial glottal stops, to determine whether the phenomenon is truly prosodic. I also use acoustic analyses to determine if voice quality is generally strengthened in prosodically-strong conditions.

### 5.2.1 Method

#### 5.2.1.1 The Boston University radio news corpus

The corpus analyzed here is the Boston University (BU) Radio News Corpus (Ostendorf, Price & Shattuck-Hufnagel, 1995). The main motivation for using the BU radio news corpus was the fact that it is labeled for prosody. Another reason was that it has been analyzed for glottalization, both vowel-initial and in all word positions, by Dilley et al. (1996) and Redi & Shattuck-Hufnagel (2001). Thus, comparison with previous work is facilitated by using the same corpus. The section of the corpus used in the present work is from the Labnews corpus, consisting of radio news read in the laboratory. The four speakers analyzed in this study form a subset of the newscasters analyzed by Dilley et al. (1996); one of their speakers (f3) was not analyzed here due to time constraints. All speakers read the same news reports. The speakers analyzed below – two female (f1a, f2b) and two male (m1b, m2b) – were adults aged 25 to 40 years old, and with no perceived regional accent.

The speech was digitized using a 16 kHz sampling rate (16 bits). Other corpus details can be found in Dilley et al. (1996). The corpus had already been labeled for prosody using

the Tones and Break Indices (ToBI) system by one or two transcribers. For the cases of two transcribers, the inter-transcriber reliability was generally high (Ostendorf et al., 1995).

In the ToBI labeling system the tone and break index tiers provide the core prosodic analysis. The tone tier in Mainstream American English (MAE)-ToBI (Beckman & Ayers Elam, 1997) consists of labels for high (H) and low (L) tones marked with diacritics indicating their intonational function as parts of pitch accents (indicated by an asterisk, e.g. H*), as phrase boundary tones, which indicate the edges of intonation phrases (indicated by a following %, e.g. H%) or as intermediate phrase accents, which indicate smaller prosodic phrasal tones (indicated by a following dash, e.g. H-).

The break index tier is used to mark the prosodic grouping of words in an utterance. The end of each word is coded for the perceived strength of its association with the next word, on a scale from 0 (for the strongest perceived juncture) to 4 (for the most disjoint). A break index of 3 usually corresponds to the end of an intermediate phrase (iP) in English, whereas a break index of 4 typically corresponds to the end of an intonation phrase (IP). In MAE-ToBI, a break index of 0 is normally used for the ends of proclitics and function words closely conjoined to the following word, and a break index of 1 for words within the same intermediate phrase. A break index of 2 is used when the perceived tone/break mismatches the perceived grouping, either because a phrase boundary is perceived in the absence of a phrase accent, or because there is a phrase accent in the absence of a perceived phrasal boundary. Because in this corpus the presence of a breath or pause following a break index of 4 was transcribed, I will refer to this as the end of an utterance within a breath group, labeled as break index 5 (cf. Price, Ostendorf, Shattuck-Hufnagel & Fong, 1991). The utterance domain above the phrasal one has sometimes been shown to exhibit greater levels of prosodic strengthening (Fougeron & Keating, 1997; Keating, Cho, Fougeron & Hsu, 2003), and thus could be relevant for the present study.

### 5.2.1.2 Coding of the BU news radio corpus for the present study

In the present study, all vowel-initial words were extracted from the corpus. A total of 2010 vowel-initial words were extracted for the four speakers, as shown in Table 5.1. This number is smaller than that analyzed by Dilley et al. (1996) because, due to time constraints, not all the paragraphs from the Labnews corpus were analyzed. In addition to word-initial vowels, 1298 word-initial sonorants (/j, w, l, ɹ, m, n/) were extracted, as well as the following vowels. For example, for a word like *Massachusetts*, the initial /m/ and following /æ/ were extracted from the corpus. Sonorant-initial words will be used to determine whether word-initial glottalization is found for all voiced sounds. In total, 1291 vowels following word-initial sonorants were extracted. This means that seven post-sonorant vowels were not extracted from the total of 1298 sonorant-initial words. These were all cases of a sonorant followed by a syllabic [ɫ] (e.g. *will* pronounced as [wɫ]), where the boundary between vowel and coda was hard to determine or did not exist.

Table 5.1: Distribution of tokens across the four speakers. Tokens with a full glottal stop [ʔ] vs. forms of voicing irregularity ([ʔ̞, ̰ ]) are indicated in parentheses.

| Speaker | Total number of tokens | Word-initial vowels | Word-initial sonorants | Vowels after word-initial sonorants |
|---------|----------|----------|----------|----------|
| f1a | 944 (126; 148) | 395 (125; 89) | 283 (1; 23) | 266 (0; 36) |
| f2b | 1281 (82; 421) | 568 (81; 285) | 356 (1; 60) | 357 (0; 76) |
| m1b | 1128 (30; 263) | 512 (30; 199) | 309 (0; 14) | 307 (0; 48) |
| m2b | 1246 (65; 316) | 535 (64; 185) | 350 (1; 37) | 361 (0; 94) |
| Total | 4599 (303; 1149) | 2010 (300; 760) | 1298 (3; 391) | 1291 (0; 254) |

The corpus was then coded by two undergraduate research assistants trained at labeling acoustic irregularity. The coders were unaware of the purpose of the study, and thus were unbiased in their coding. In addition to coding for presence and type of irregularity (to be described below), the coders also transcribed the presence of a variety of other factors,

described in further detail below. I then reviewed the corpus data and arbitrated between-coder differences. The agreement rate for codings of irregularity was over 90%.

Generally, the coding of voicing irregularity followed that described by Dilley et al. (1996) and Redi & Shattuck-Hufnagel (2001). First, the coders rated whether there was a percept of 'glottalization', regardless of whether the percept was due to a glottal stop. Tokens with weak percepts of glottalization with no clear acoustic evidence for voicing irregularity were labeled as glottalized, unlike in Dilley et al. (1996), where such tokens were excluded from the analysis. I included these tokens for the purposes of the voice quality analysis (described below). Such tokens represented only 2% of total words in the corpus, and therefore were unlikely to have a significant influence on the subsequent analysis.

Second, if there was a percept of glottalization, the coders labeled the *type* of aperiodicity found, based on inspection of the waveform. This labeling provides visual support for the percept of glottalization, but the individual types of aperiodicity will not be analyzed. Four types were identified, three of which (aperiodicity, diplophonia, and creak) followed the description by Redi & Shattuck-Hufnagel (2001). *Aperiodicity* is defined as pulse-to-pulse irregularity, either as jitter or as visible noise. *Diplophonia* refers to irregularity characterized by regular alternation in shape, duration, or amplitude of glottal periods. Thus, for diplophonia the pulse-to-pulse alternation is sustained, in contrast to the sudden, unpredictable changes to pulse shape found for 'aperiodicity.' *Creak* refers to low F0 accompanied by near-total damping of glottal pulses, commonly (but not exclusively) found phrase-finally. Redi & Shattuck-Hufnagel (2001) identify another type of irregularity which they term *glottal squeak*, but such cases were not found in this corpus, probably because in that study the authors identified cases of irregularity occurring anywhere in a word, not just word-initially. Together, aperiodicity, diplophonia, creak (and squeak) represent the cases of *voicing irregularity*.

In this study, a *full glottal stop* was also identified. In the corpus, [ʔ] only occurred word-initially; no cases of [ʔ] as an allophone of /t/ were extracted. Thus, [ʔ] was characterized by a period of silence of at least two pulses, followed by a burst and subsequent onset of

88

phonation (due to the following vowel) which becomes increasingly modal. If preceded by a voiced sound, the glottal stop often showed an impulse (visually distinct from the pulses belonging to the preceding voiced sound) right before the glottal closure (evidenced by the absence of a signal in the waveform). This likely corresponds to the energy produced by the abrupt closure of the vocal folds. An example of [ʔ] is shown in Figure 5.1.
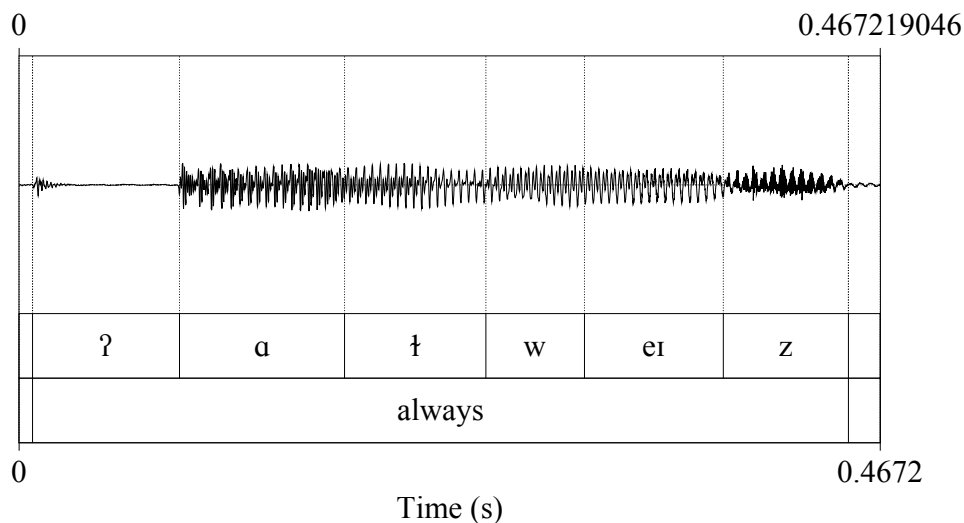


Figure 5.1: Example of a glottal stop at the onset of 'always,' uttered by speaker f1a.

For a vowel-vowel sequence with creak between the two vowels, it can be difficult in principle to determine whether any of the creaky pulses are in fact the burst of a glottal stop. However, in practice such difficulty differentiating creak from full glottal stops rarely arose. Pulses during creak, though irregular in period, were not separated by more than a two-pulse period of silence. Thus, there was rarely a debate whether a sequence of two vowels corresponded to [v̰#v] vs. [v̰#ʔv]. There were only ten such problematic cases, which were labeled conservatively as just having creak.

Different types of irregularity were sometimes found for the same segment, and so multiple types could be coded per token. For example, diplophonia was sometimes found during intervals of creak. Additionally, aperiodic and/or creak-like phonation was sometimes found after a glottal stop. For example, in Figure 5.1, the vowel following the glottal stop begins with aperiodicity.

89

### 5.2.1.3 Other factors in the analysis

In addition to coding for presence and type of glottalization, the coders also recorded prosodic, lexical, and segmental information. The prosodic factors are summarized in Table 5.2. The factor 'prominence' refers to a syllable with prosodic prominence, either due to the presence of a pitch accent, or if the syllable belonged to a function word, but had an unreduced vowel (e.g. [ænd] for *and*), or both. Thus, prominence represents a superset of pitch-accented syllables, and is more like the traditional 'stress,' where full vowels are said to bear secondary stress. The reason for including this factor was that some vowels had perceived prominence, but no pitch accent was marked in the BU Corpus. Further inspection sometimes revealed a pitch excursion indicative of a potential pitch accent despite none being coded, but usually the vowel was unreduced, which is unexpected for function words. The absence of expected vowel reduction is not typically used as a cue for vowel prominence, but the coders agreed that these words were more prominent than expected. Often these words occurred phrase-initially, suggesting that the absence of vowel reduction is related to phrase-initial strengthening (Cho & Keating, 2009). Thus, in phrase-initial position, vowels – even when not pitch-accented – are nonetheless more strongly articulated and thus more perceptually prominent than they would be phrase-medially.

Aside from prosodic factors, lexical and segmental factors were also included, and they are summarized in Table 5.3. Lexical frequencies were taken from the SUBTLEX$_{WF}$-US corpus, whose lexical frequencies are thought to be more representative of currently-spoken English than are those from older corpora (Brysbaert & New, 2009).

### 5.2.1.4 Acoustic measures

Beyond the presence of a full [ʔ], I also obtained quantitative data from acoustic measures to provide a gradient analysis of the strength of glottalization. These measures can also help

Table 5.2: Prosodic factors in the analysis.

| Factor | Explanation |
| --- | --- |
| Preceding break index | Break index (from 0-5) between target word and prec. word |
| Following break index | Break index (from 0-5) between target word and foll. word |
| Pitch accent | Presence of pitch accent on target syllable |
| Pitch accent type | Type of pitch accent on target syllable (H*, L*, etc.) |
| Prominence | Presence of a pitch accent and/or unreduced, stressed vowel |
| Boundary tone | Presence of boundary tone/phrase accent on target syllable |
| Boundary tone type | Type of boundary tone/phrase accent on target syllable |
| Preceding pause | Presence of a pause before target syllable |
| Preceding glot. | Presence of glottalization-like irregularity on preceding syllable |

Table 5.3: Segmental and lexical factors in the analysis.

| Factor | Explanation |
| --- | --- |
| Vowel | The type of vowel in the target syllable |
| Vowel height | Whether the target vowel was high, mid, or low |
| Vowel frontness | Whether the target vowel was front, central, or back |
| Vowel length | Whether the target vowel was tense or lax |
| Word | The word containing the target syllable |
| Log frequency of word | Log frequency of target word |
| Word type | Whether target word a content or function word |
| Preceding sound | Final sound of preceding word |
| Hiatus | Potential for hiatus (i.e. prec. sound was a vowel) |
| Vowel quality of prec. vowel | Height, frontness, and length of prec. vowel. |
| Prec. word | The word preceding the target syllable |
| Log freq. of prec. word | Log frequency of the prec. word |
| Prec. word type | Whether prec. word was content or function word |

determine which cases of voicing irregularity are in fact lenited stops [ʔ] vs. phrasal creak, provided the two differ in their acoustic realization. The acoustic measures included in the analysis, along with their relation to voice quality, are described in Table 5.4. To obtain the measures, the coders manually segmented the word-initial vowels in the corpus. (Although segment boundaries had already been provided in the corpus, many had been aligned automatically, and many files had not been checked for segment boundaries). VoiceSauce (Shue et al., 2011) was then run over the entire sound file (not just the labeled vowels), because many tokens were so short that they required longer windows of analysis in order to ob-

tain acoustic measures. The acoustic measures were then averaged over the entire vowel's duration.

If voicing irregularity in the corpus is only due to increased constriction (i.e., a glottal stop target), the spectral and noise measures listed in Table 5.4 are expected to be lower than for vowels with no voicing irregularity. In addition, laryngealized phonation often involves a decrease in fundamental frequency or F0. Crucially, voicing with increased constriction typically shows lower values of H1*-H2* (Garellek & Keating, 2011), and lower values of the measure (either corrected or uncorrected for vowel formants) are correlated with higher values of flow adduction quotient (Holmberg et al., 1995), increased values of EGG contact quotient (DiCanio, 2009; Kuang, 2011; Esposito, 2012), and lower open quotient derived from glottal area (Shue, Chen & Alwan, 2010a; Kreiman, Shue, Chen, Iseli, Gerratt, Neubauer & Alwan, 2012). Therefore, H1*-H2* is taken to be the likeliest acoustic measure to be correlated with increased glottal closure.

### 5.2.2 Results

Of the 2010 vowel-initial words extracted from the corpus, 1060 or 53% showed at least one form of irregularity. Only 300 or about 15% of all word-initial vowels had full glottal stops. Vowel-initial words showing a form of voicing irregularity with no glottal stop accounted for 37% of all vowel-initial words and about 72% of cases of irregular word-initial vowels (vowels with a glottal stop, aperiodicity, diplophonia, and/or creak). Not surprisingly, none of the vowels after sonorants (e.g., the /æ/ in *Massachusetts*) had glottal stops, but about 20% showed voicing irregularity. 30% of the sonorants had irregular voicing, and only three cases of glottal stops before sonorants were documented. This number is negligible, and as these cases are likely instances of creak with a long lag between the first pulse and the next, they will not be discussed further. Across speakers, the rates of voicing irregularity and glottal stops were 27% of all tokens for f1a, 39% for f2b, 26% for m1b, and 31% for m2b.

92

Table 5.4: Acoustic measures in the analysis. Asterisks indicate measures corrected for formants.

| Measure | Explanation | Relation to voice quality |
|---|---|---|
| F0 | Fundamental frequency | Pitch, correlated with prosodic tones and stress |
| Duration | Length of vowel | Correlated with prominence, prosodic position (Cole, Mo & Hasegawa-Johnson, 2010) |
| H1*-H2* | Difference between amplitudes of first two harmonics | Thought to be positively correlated with open quotient (OQ) (Holmberg et al., 1995) |
| H2*-H4* | Difference between amplitudes of second and fourth harmonics | Thought to be correlated with vocal fold stiffness (Zhang, Kreiman & Gerratt, 2011), and used in the perception of breathiness (Kreiman, Garellek & Esposito, 2011) |
| H1*-A1* | Difference in amplitudes of first harmonic and harmonic nearest F1 | Correlated with breathiness, thought to be related to presence of a posterior gap (Hanson et al., 2001) |
| H1*-A2* H1*-A3* | Difference in amplitudes of first harmonic and harmonic nearest F2, F3 | Correlated with overall spectral tilt, perhaps due to abruptness of closure (Stevens, 1977; Hanson et al., 2001) |
| CPP | Noise measure | Correlated with modal vs. non-modal voice (Garellek & Keating, 2011) |
| HNR | Noise measure (in four spectral bands) | Correlated with modal vs. non-modal voice (Garellek & Keating, 2011) |
| Energy | Measure of loudness | Correlated with prominence (Kochanski, Grabe, Coleman & Rosner, 2005) |

The distribution of [ʔ] and other forms of glottalization is shown in Table 5.1 (p. 87), and the proportion of each type of irregularity for initial vowels, sonorants, and post-sonorant vowels is shown in Figure 5.2. The glottalization rates are similar to those found by Dilley et al. (1996) in their analysis of the same corpus, though they did not look specifically at cases of full [ʔ]. The rate of full [ʔ] occurrence is larger here than what was found for two British English speakers by Bissiri & Volín (2010), but smaller than was found in German by Pompino-Marschall & Żygis (2011).
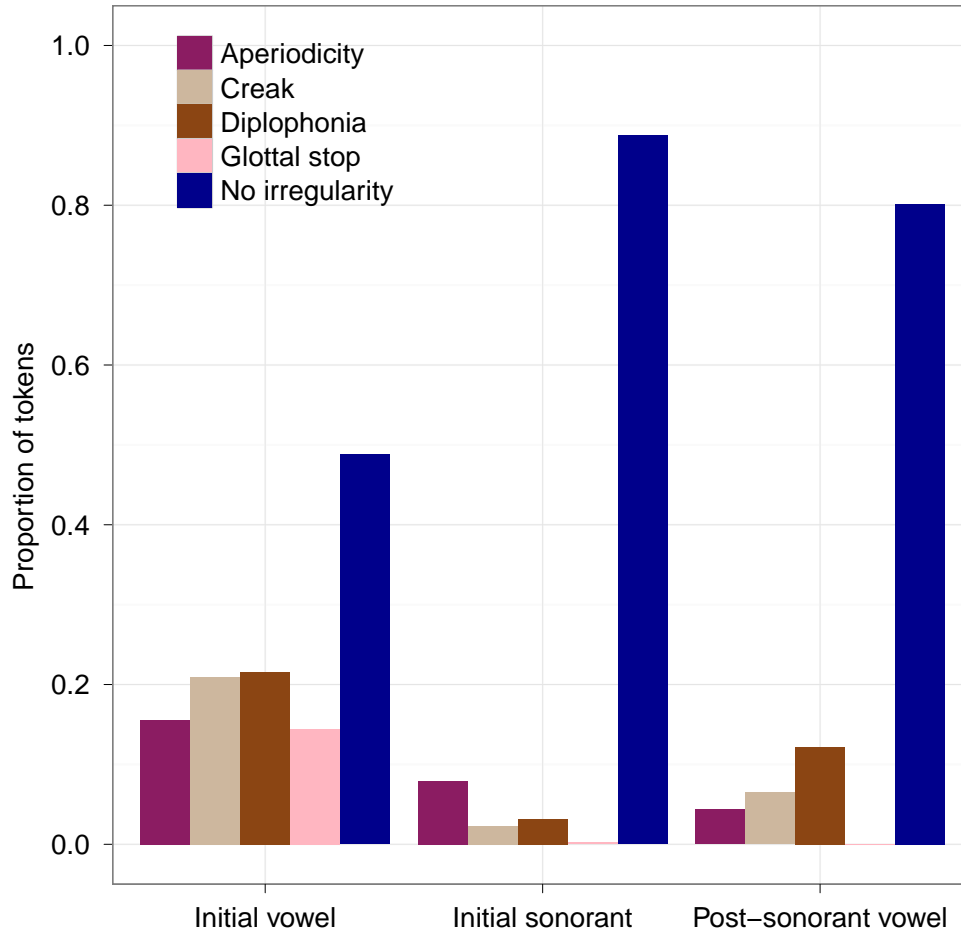
Figure 5.2: Proportion of each type of irregularity for word-initial vowels, word-initial sonorants, and post-sonorant vowels. More than one type of irregularity can be present on a given token, so the sum of irregularity types can exceed 1.

#### 5.2.2.1 Predicting full glottal stop occurrence

To predict where full glottal stops [ʔ] occur, the data were first subset into cases of word-initial vowels with a full [ʔ], and cases of no perceived/visual voicing irregularity, leaving aside cases of voicing irregularity that were not instance of full glottal stops. A mixed-effects logistic regression model was fitted to these data using the *lmer()* function in the *lme4* package (Bates, Maechler & Dai, 2008) in *R* (R Development Core Team, 2011), following Baayen (2008). The model's dependent variable was presence of [ʔ] vs. no perceived/visual voicing irregularity, and had 14 independent variables from the factors listed in Tables 5.2

and 5.3 above: previous break, pitch accent, prominence, hiatus, following break, word type, vowel height, length, and backness, presence of preceding pause and glottalization, word frequency, preceding word frequency, and preceding word type. An interaction term (presence of hiatus:preceding glottalization) was included because it improved the model's fit (which was assessed by the *anova* function in *R*, following Baayen (2008)). Speaker and word were included as random intercepts. The results are shown in Table 5.5. The coefficient estimates indicate the direction of significance, with a positive coefficient indicating an increase in the odds of there being a full glottal stop. Both an increase in preceding break index and presence of prominence on the following vowel increased the likelihood of glottal stop occurrence, and these factors were the most significant in the model. The effects of phrasal domain and prominence can be seen in Figure 5.3. Prominent vowels were more likely to be preceded by [ʔ], regardless of the preceding break. But the phrasal domain was also significant, with rates of glottal stop occurrence decreasing with a decrease in preceding break index.

Other significant predictors included presence of a preceding pause and preceding 'glottalization,' both of which increased the likelihood of obtaining a full [ʔ]. A preceding pause might increase the likelihood of a glottal stop for two reasons. First, pauses had already been marked in the corpus, but it was apparent to the coders that some of these represented the closure durations for glottal stops rather than true pauses. Second, true pauses increase the dissociation between two words, such that a break index of 4 with no pause is weaker than a 4 followed by a pause. Preceding glottalization (mostly from phrase-final creak) might increase the likelihood of there being a following glottal stop because the vocal folds are mostly abducted and closing irregularly during creak (Slifka, 2006). Thus, vocal fold closure for [ʔ] could help resume phonation after a period of creak if the vocal folds are vibrating irregularly. There was a significant interaction between preceding glottalization and vowel-vowel hiatus, as shown in Figure 5.4. A hiatus environment (i.e. a vowel-initial word that was preceded by a word ending in a vowel) was found to be a significant predictor of full [ʔ] only when the preceding word ended in glottalization (i.e., with some form of irregularity). One could also

95

Table 5.5: Results of logistic regression model predicting occurrence of [ʔ] vs. no glottalization for vowel-initial words.

| | Coef $\beta$ | SE($\beta$) | z | p |
|---|---|---|---|---|
| Intercept | −6.15 | 1.24 | −5.0 | **<.0001** |
| Preceding break | 1.19 | 0.14 | 8.4 | **<.0001** |
| Hiatus=Y | −0.06 | 0.41 | −0.1 | >0.9 |
| Accent=Y | 0.41 | 0.31 | 1.3 | >0.2 |
| Prominence=Y | 4.03 | 0.38 | 10.6 | **<.0001** |
| Preceding glottalization=Y | 1.26 | 0.32 | 3.9 | **<.0001** |
| Following break | 0.24 | 0.13 | 1.8 | >0.1 |
| Word type=function | −0.03 | 0.60 | 0.0 | >1 |
| Vowel frontness=front | 0.26 | 0.64 | 0.4 | >0.7 |
| Vowel frontness=central | 1.00 | 0.76 | 1.3 | >0.2 |
| Vowel height=low | 0.57 | 0.52 | 1.1 | >0.3 |
| Vowel height=mid | −0.08 | 0.56 | −0.1 | >0.9 |
| Vowel length=lax | 0.06 | 0.54 | 0.1 | >0.9 |
| Preceding pause=Y | 2.12 | 0.40 | 5.3 | **<.0001** |
| Log freq. word | −0.22 | 0.19 | −1.1 | >0.3 |
| Log freq. preceding word | −0.30 | 0.16 | −1.9 | >0.1 |
| Preceding word type=function | 1.15 | 0.47 | 2.4 | **<.05** |
| Hiatus=Y:Preceding glottalization=Y | 1.96 | 0.75 | 2.6 | **<.01** |

assume that glottal stops occurred more often when there was preceding voicing irregularity because the glottal stops triggered anticipatory laryngealization. That is, instead of being due to phrase-final creak independent of a following glottal stop, the preceding irregularity would instead be caused by the following glottal stop. This does not appear to be the case, however, because of the words beginning with glottal stops, nearly 73% were *not* preceded by voicing irregularity. This could be due to the fact that the preceding word was not voiced, so I also checked to see whether in hiatus environments, the word-initial vowel could have a [ʔ] but no preceding irregularity before the glottal stop. The answer is yes: of words beginning with [ʔ] and in hiatus environments, 49% do not have previous 'glottalization.' This strongly implies that the cause of the preceding irregularity is independent of the following [ʔ].

The final predictor to emerge as significant from the model (though much less so than prominence or preceding break) was the preceding word type. A preceding function word (compared to a content word) increased the likelihood of a word-initial vowel's being preceded
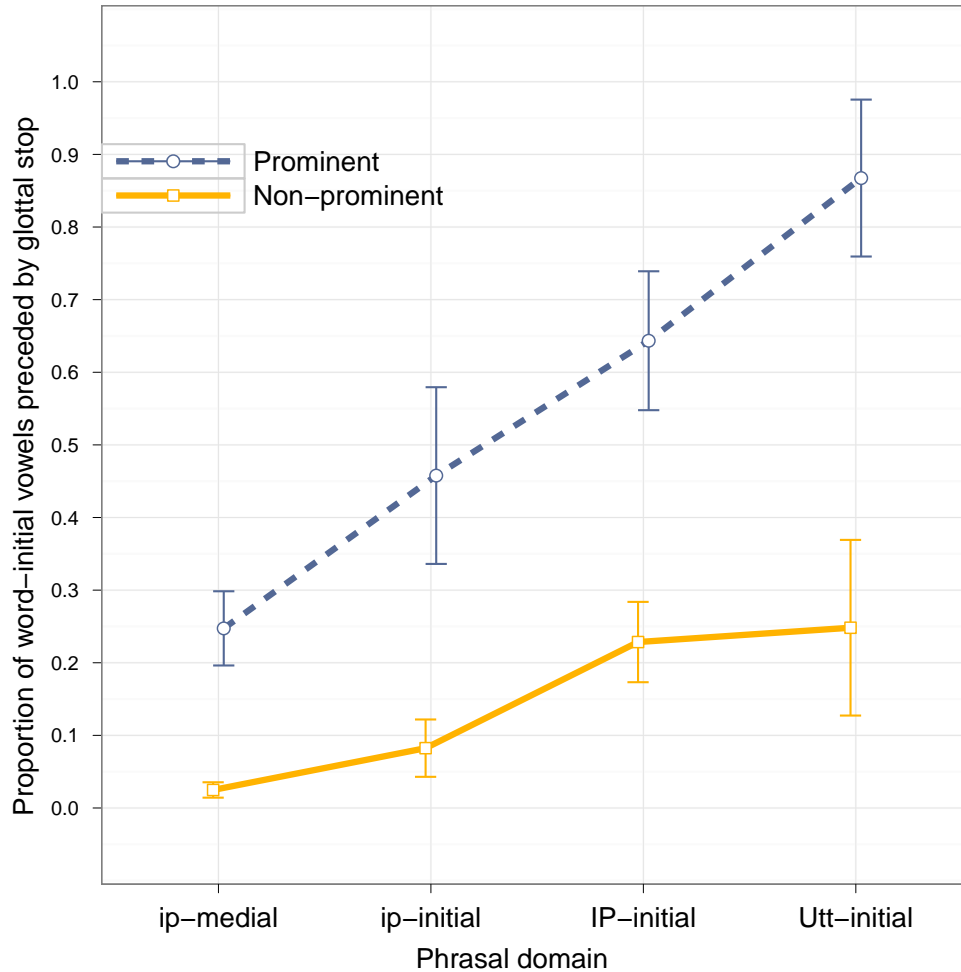
Figure 5.3: Proportion [ʔ] for vowel-initial words as a function of phrasal domain and prominence. Error bars indicate 95% confidence intervals.

by a full glottal stop, possibly because a glottal stop will render the vowel-initial word more prominent by preventing the function word from becoming a proclitic on the target word. For example, the sequence 'the only' is likely to be pronounced without a clear boundary between the determiner and the adjective (e.g., as [ðəoʊnli] or even [ðioʊnli]). If produced with a full glottal stop ([ðəʔoʊnli]), the boundary between the determiner and the adjective is clearly defined, which might increase the degree of perceived prominence of the content word to listeners because the function word has not become a proclitic.

The relative importance of each of the significant factors was also assessed by comparing the full model to smaller models, each lacking one of the significant factors. This form of
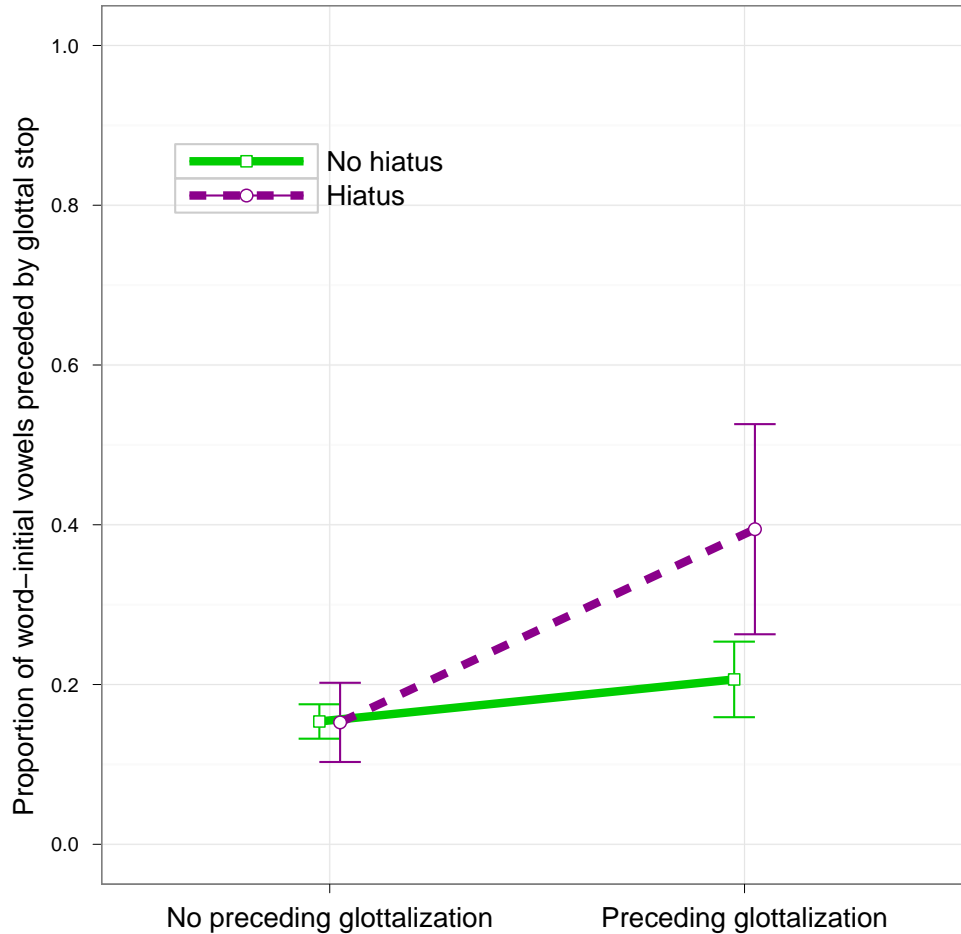
Figure 5.4: Interaction between vowel hiatus and preceding glottalization in the occurrence of [ʔ]. Error bars indicate 95% confidence intervals.

model comparison, done by means of the *anova* function in $R$, provides a chi-squared statistic and $p$-value indicating whether the full model provides a significantly better fit to the data than the model with a factor removed (Baayen, 2008). The results mirror the $z$-scores of the estimates in the full model, indicating that the most important factors are, in order, prominence > preceding break > preceding glottalization > preceding pause > hiatus > preceding word type.

In sum, full glottal stops are more likely to occur when the vowel-initial word is phrase-initial and when the vowel is prominent. Preceding pauses or glottalization, hiatus, and the preceding word type were also found to be significant predictors of full glottal stops, but

much less so than prominence and preceding break index. By considering only the effects of prominence and phrasal position, it is possible to account for 95% of cases of [ʔ], as shown in Figure 5.5. Prominence alone is able to account for three quarters of cases, and phrase-initial position for nearly seven of ten cases.
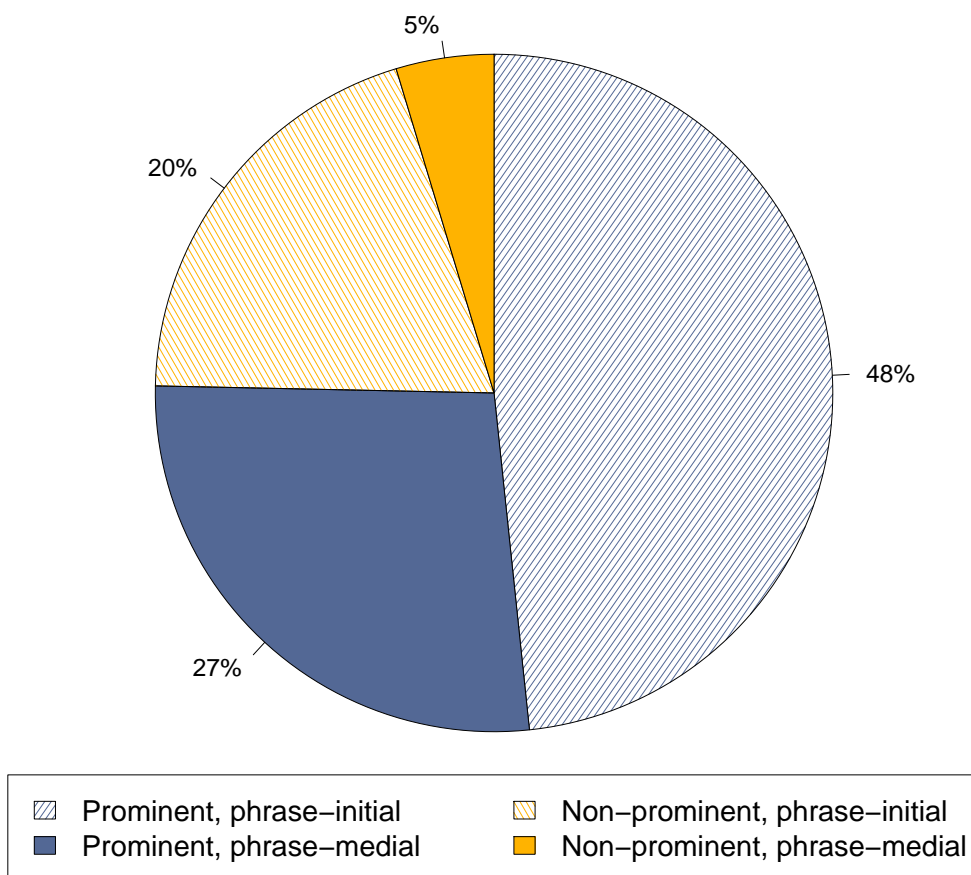


Figure 5.5: Distribution of full glottal stops [ʔ] as a function of prominence and phrasal position. Prominence (violet dashed and solid slices) accounts for 75% of the occurrences [ʔ], phrase-initial position (dashed slices) accounts for 68% of occurrences. 95% of full glottal stops can be attributed either to prominence, or phrase-initial position, or to both.

### 5.2.2.2 Voice quality of vowels following [ʔ]

Although it is not possible to obtain acoustic measures of voice quality for a (voiceless) [ʔ], the voice quality of the following vowel can be investigated. To do so, I ran a logistic regression model predicting [ʔ] vs. no glottalization to determine which acoustic measures differentiate vowels following glottal stops from sounds with no visual/auditory cues to glottalization. A logistic mixed-effects model was fitted to the data, with [ʔ] vs. no glottalization as the dependent variable, the 13 acoustic measures (listed in Table 5.4) as fixed effects, and item, sound, and speaker as random intercepts. In addition, a random slope of duration by speaker was included because it improved the model's fit. No interaction terms between any acoustic measures significantly improved model fit. The results are shown in Table 5.6. Many acoustic measures differentiated vowels following [ʔ] from other vowels. The most important factor was duration, which is longer for vowels following [ʔ] than for vowels with no glottalization. This is probably an effect of prominence, given that duration is a known correlate of prominence (see Turk & Sawusch (1996); Fant, Kruckenberg & Liljencrants (2000); Cole et al. (2010), and references therein), or of phrasal position. HNR <1500Hz, H1*-A2*, and H1*-H2* are significantly lower for vowels following [ʔ], consistent with the idea that vowels after a glottal stop are laryngealized. Interestingly, HNR <2500 Hz was significantly higher for these vowels, which must be due to a boost in harmonic energy in the frequencies between 1500 Hz and 2500 Hz. The abrupt closure of the vocal folds during laryngealization is known to increase energy in the higher frequencies (Kreiman & Sidtis, 2011), and these results imply that the energy boost is within the 1500 Hz to 2500 Hz range.

Another surprising finding is that H1*-A1* was higher for vowels following [ʔ]. Higher values of H1*-A1* might relate to posterior opening of the cartilaginous glottis, with higher values of the measure correlated with larger posterior gaps and thus breathiness (Hanson et al., 2001). Activation of the vocal fold abductor muscle (the PCA) is known to occur before the release of a hard glottal attack, and this activation forces the arytenoids apart (Hirose & Gay, 1973), perhaps causing H1*-A1* to rise.

100

Finally, the decrease in energy in vowels following [ʔ] can be attributed to laryngealization following the release of the glottal stop (as suggested by the lower values of H1*-H2* and H1*-A2*). Therefore, although most of these vowels following [ʔ] are prominent, and loudness is a cue to prominence (Kochanski et al., 2005, but cf. Turk & Sawusch, 1996), vowels following [ʔ] are not louder, probably because of the laryngealized voice quality, which often shows a decrease in energy (Gordon & Ladefoged, 2001).

Thus, vowels following [ʔ] show acoustic characteristics typical of laryngealized phonation produced with increased glottal closure and aperiodicity. These effects are strong enough to affect the voice quality measures after they have been averaged over the entire vowel's duration. One notable exception is that H1*-A1* is higher for these vowels, which could be due to abduction of the arytenoids necessary to resume phonation after a glottal stop.

Knowing now that the vast majority of full glottal stops occur in prominent and/or phrase-initial environments, I turn to the cases of word-initial vowels *without* full glottal stops to determine if they show voice quality that is characteristic of laryngealization, i.e. voicing with more vocal fold closure. If they do, I will infer that such laryngealization in prominent and/or phase-initial environments is due to incomplete glottal stops ([ʔ̞]), because these same environments are known to be the most important factors in predicting full glottal stops.

### 5.2.2.3   Acoustic effects of prominence vs. phrasal strengthening on word-initial vowels

If prosody and phrasing are the most important predictors of full [ʔ], I hypothesized that they should also be good predictors of incomplete [ʔ̞] as well. To test this, I looked at the voice quality of vowels without a full [ʔ], to see if they show characteristics of laryngealization, which would be consistent with the presence of an incomplete glottal stop. Note that by

Table 5.6: Significance of the fixed effects in the logistic model predicting vowels following [ʔ] vs. initial vowels with no glottalization from the acoustic measures.

| | Coef $\beta$ | SE($\beta$) | z | p |
|---|---|---|---|---|
| Intercept | −6.65 | 0.59 | −11.2 | <.0001 |
| H1*-H2* | −0.11 | 0.05 | −2.1 | <.05 |
| H2*-H4* | 0.05 | 0.04 | 1.3 | >0.2 |
| H1*-A1* | 0.14 | 0.05 | 3.1 | <.01 |
| H1*-A2* | −0.22 | 0.04 | −5.5 | <.0001 |
| H1*-A3* | 0.04 | 0.02 | 1.7 | >0.1 |
| F0 | 0.02 | 0.00 | 4.1 | <.0001 |
| Duration | 0.04 | 0.01 | 7.5 | <.0001 |
| HNR < 500Hz | −0.02 | 0.05 | −0.4 | >0.7 |
| HNR < 1500Hz | −0.44 | 0.08 | −5.6 | <.0001 |
| HNR < 2500Hz | 0.33 | 0.13 | 2.5 | <.05 |
| HNR < 3500Hz | −0.17 | 0.10 | −1.7 | >0.1 |
| CPP | −0.27 | 0.08 | −3.4 | <.001 |
| Energy | −0.36 | 0.13 | −2.8 | <.01 |

'laryngealization' I refer specifically to voicing with increased glottal closure (Gordon & Ladefoged, 2001; Esling et al., 2005).

Recall from Section 5.2.1.4 that H1*-H2* is taken to be the likeliest acoustic measure to be correlated with the proportion of the glottal cycle during which the vocal folds are closed (i.e., longer closure = more glottal-stop-like voicing). Indeed, lower values of H1*-H2* are found for vowels following full [ʔ] (see Table 5.6). To test if lower values of H1*-H2* are associated with prominence on vowel-initial words that are not preceded by [ʔ], I fitted a linear mixed-effects model predicting H1*-H2* as a function of the prominence and phrasing. H1*-H2* (standardized and with outliers with an absolute $Z$-score > 3 removed) is the dependent variable, and phrasal condition*prominence (main effects and interaction) are the fixed effects. Vowels preceded by a break index (BI) of 0 or 1 were recoded as 'ip-medial,' those preceded by BI 3 and 4 as (respectively) 'ip-initial' and 'IP-initial,' and those preceded by BI 5 as 'Utterance-initial.' I excluded vowels preceded by a break index of 2, because these are unclear cases that could be phrase-medial or phrase-initial. Random intercepts are included for speaker and word, as well as a random slope of mean F0 by

speaker. The prosodic condition factor is first coded using forward difference coding, such that (keeping prominence constant), the mean of each prosodic condition is compared to the mean of the following (higher) condition. Additionally, the difference in prominence across all prosodic conditions ('main effect' of prominence) is assessed separately. Pairwise comparisons between prominent vs. non-prominent sounds at a given prosodic domain are assessed by changing the contrast coding of the prosodic condition to zero-sum coding and subsequent reference changes. Significance of a pairwise difference in means is evaluated based on whether the absolute $t$-value was greater than 2, given that MCMC sampling has not yet been implemented for models with random slopes (Baayen, 2008). Phrasal condition is also recoded as a linear factor (by increasing break index), in order to assess its main effect within prominence category. The results for word-initial vowels are shown in Figure 5.6. H1*-H2* is plotted as a function of prominence and phrasal domain. No difference as a function of prosodic domain is found. The difference in H1*-H2* as a function of vowel prominence is significant for ip-medial and IP-initial vowels. The main effect of prominence on H1*-H2* (across all phrasal domains) is also significant. $T$-values can be found in Table 5.7.

Table 5.7: $T$-values for pairwise comparisons of (standardized) H1*-H2* within and across prominence groups for word-initial vowels not preceded by [?] (see also Figure 5.6). Values below -2 or greater than 2 are considered significant.

(a) Difference in H1*-H2* within prominence.

|  | ip-medial vs. ip-initial | ip-initial vs. IP-initial | IP-initial vs. Utt-initial | Main effect-phrasing |
|---|---|---|---|---|
| Prominent | -0.62 | 0.77 | -0.12 | 0.12 |
| Non-prominent | 0.29 | -1.51 | -0.29 | -1.61 |

(b) Difference in H1*-H2* for prominent vs. non-prominent initial vowels.

| ip-medial | ip-initial | IP-initial | Utt-initial | Main effect of prominence |
|---|---|---|---|---|
| **-4.04**\* | -1.04 | **-2.80**\* | -1.08 | **-2.78**\* |

These results are consistent with the idea that prominent word-initial vowels – even those with no [?] – are produced with increased vocal fold constriction. Surprisingly, higher
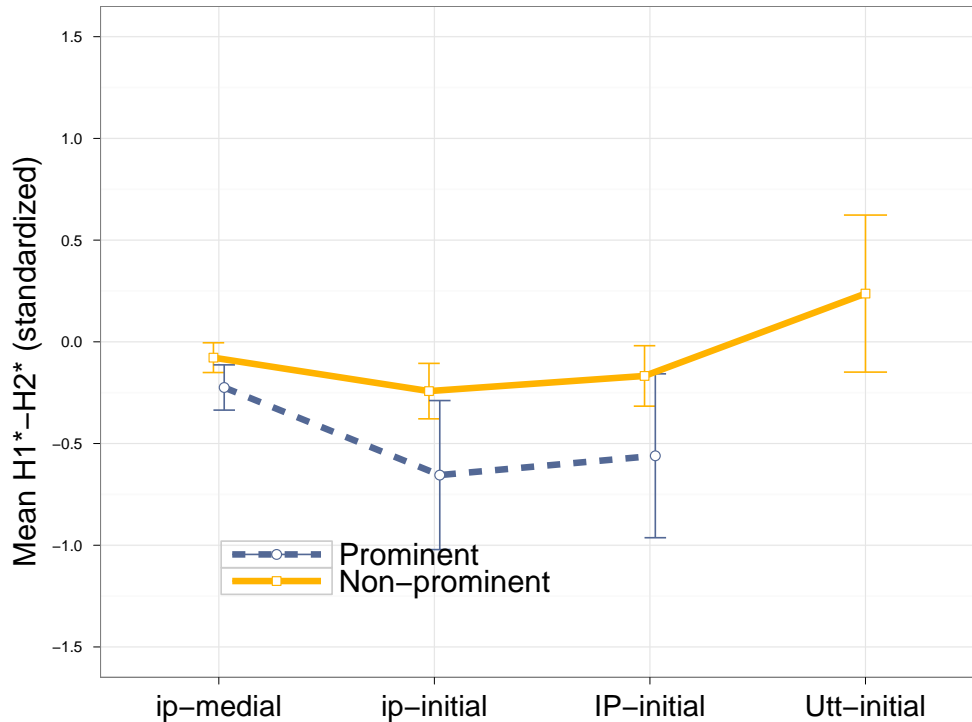
Figure 5.6: Mean H1*-H2* for word-initial vowels. Prominent vowels are the dashed lines. Error bars indicate 95% confidence intervals. The mean for Utterance-initial non-prominent vowels is removed because only 5 tokens were *not* preceded by [ʔ].

prosodic domains are not associated with a decrease in H1*-H2*. This is inconsistent with the assumption that larger prosodic disjunctures trigger an increase in glottal stops. Therefore, for word-initial vowels that are not preceded by [ʔ], prominence induces greater laryngealization (based on lower values of H1*-H2*), whereas higher prosodic domains do not.

### 5.2.2.4 Acoustic effects of prominence vs. phrasal strengthening on word-initial sonorants and their following vowels

If the lowering of H1*-H2* for prominent vowel-initial words is due to a glottal stop gesture (which can be completely or incompletely realized), then we would expect that such lowering would not occur for word-initial sonorants or their following vowels, because these positions are never preceded by a glottal stop in English. To test this, I fitted linear mixed-effects models to word-initial sonorants and then to their following vowels, in order to predict H1*-

H2* or H1-H2 as a function of the prominence and phrasing. This was done using models identical in structure to those fitted to the word-initial vowels. However, for word-initial sonorants, H1-H2 (uncorrected for vowel formants) was used as the dependent variable, because formant tracking errors during the sonorants affected the values of (corrected) H1*-H2* (Garellek, 2012).

The overall results for word-initial sonorants are shown in Figure 5.7a and Table 5.8). No significant main effect of prominence is found, but the main effect of phrasing is significant for both prominent and non-prominent sonorants. Higher phrasal domains are generally associated with an increase in H1-H2, even if pairwise differences between domains are not always significant. Because the effect of formants on H1-H2 will differ based on the type of sonorant, the H1-H2 results are also separated according to the different sonorants in Figure D.1 in Appendix D. The overall pattern is generally found for /j, w, l, ɹ/, but not for nasals, where there is little effect of phrasing or prominence on H1-H2.

The results for vowels after word-initial sonorants are shown in Figure 5.7b. No significant main effects or interactions were found (see Table 5.9). Therefore, the overall results show that, unlike for word-initial vowels, H1*-H2*/H1-H2 is not lower for prominent initial sonorants or post-sonorant vowels. This is consistent with the hypothesis that only word-initial vowels should show laryngealization, because only they are preceded by a glottal stop gesture (potential reasons for this will be discussed in Section 5.4). Crucially, the increase in H1-H2 as a function of phrasing for word-initial sonorants suggests that phrase-initial voicing is generally breathier, not more laryngealized.

### 5.2.3  Discussion of corpus study

This study seeks to answer two questions regarding English word-initial glottal stops:

1. Which factors matter most in predicting the occurrence of glottal stops?

2. Is word-initial glottalization due to voice quality strengthening?

(a) Initial sonorants        (b) Post-sonorant vowels
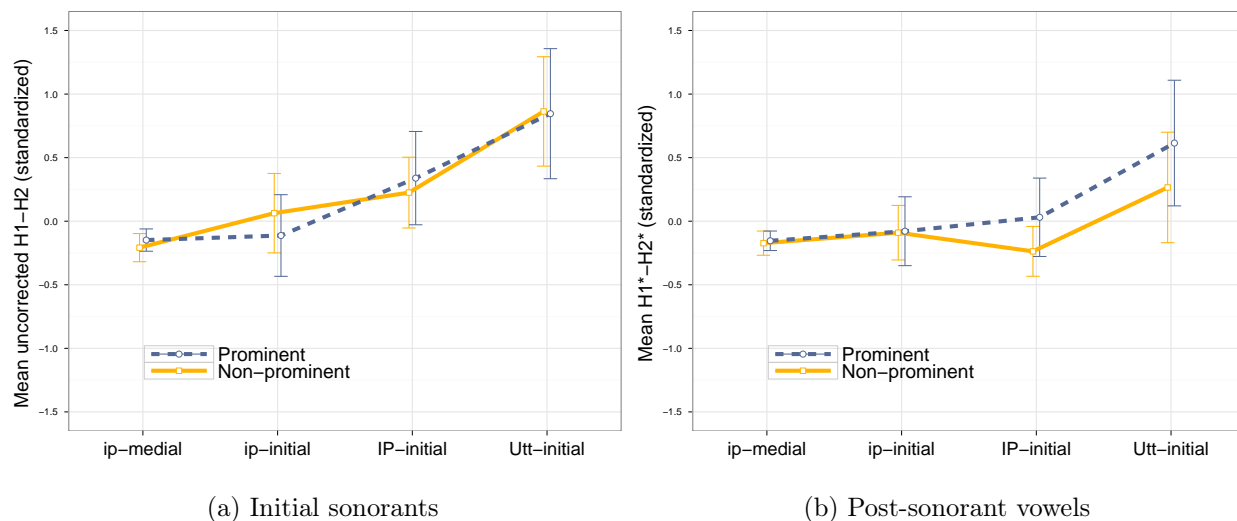
Figure 5.7: Mean H1*-H2* (or H1-H2) for word-initial sonorants (a), and post-sonorant vowels (b). H1-H2 is used for initial sonorants because of problems with formant tracking that affect corrected measures (Garellek, 2012). Prominent vowels are the dashed lines. Error bars indicate 95% confidence intervals.

Table 5.8: *T*-values for pairwise comparisons of (standardized) uncorrected H1-H2 within and across prominence groups for word-initial sonorants (see also Figure 5.7a). Values below -2 or greater than 2 are considered significant.

(a) Difference in H1-H2 within prominence.

|  | ip-medial vs. ip-initial | ip-initial vs. IP-initial | IP-initial vs. Utt-initial | Main effect-phrasing |
|---|---|---|---|---|
| Prominent | -0.17 | **2.62**\* | 1.23 | **5.09**\* |
| Non-prominent | **3.35**\* | 1.84 | **2.76**\* | **3.59**\* |

(b) Difference in H1*-H2* for prominent vs. non-prominent initial sonorants.

| ip-medial | ip-initial | IP-initial | Utt-initial | Main effect of prominence |
|---|---|---|---|---|
| -0.57 | 0.88 | -0.34 | -1.59 | -1.91 |

The results from Section 5.2.2.1 show that [ʔ] is predicted largely by prominence and phrase-initial position, which together can account for 95% of cases of full glottal stops. This finding is in line with previous research, which has shown that prominence (accent or stress) and phrasing are important in predicting glottal stops and/or glottalization (Pierrehumbert & Talkin, 1992; Pierrehumbert, 1995; Dilley et al., 1996; Redi & Shattuck-Hufnagel, 2001; Davidson & Erker, 2012).

Table 5.9:  *T*-values for pairwise comparisons of (standardized) H1\*-H2\* within and across prominence groups for post-sonorant vowels (see also Figure 5.7b). Values below -2 or greater than 2 are considered significant.

(a) Difference in H1-H2 within prominence.

|  | ip-medial vs. ip-initial | ip-initial vs. IP-initial | IP-initial vs. Utt-initial | Main effect-phrasing |
|---|---|---|---|---|
| Prominent | 0.03 | -0.58 | -0.66 | 0.51 |
| Non-prominent | -1.34 | 0.57 | -0.95 | 1.50 |

(b) Difference in H1\*-H2\* for prominent vs. non-prominent post-sonorant vowels.

| ip-medial | ip-initial | IP-initial | Utt-initial | Main effect of prominence |
|---|---|---|---|---|
| 0.07 | 1.04 | -0.05 | 0.10 | -0.57 |

To find cases of glottal stops that have been realized incompletely, in Section 5.2.2.3 I looked at the effects of prominence and phrasal position on the acoustic properties of word-initial vowels. I focused on these two factors because they are by far the most important factors in predicting full [ʔ] occurrence, and are therefore also likely to predict occurrences of incomplete [ʔ̰]. The results show that prominence can in fact be used to predict laryngeal-ization that is typical of voicing with increased glottal closure. However, this is only true for word-initial vowels. This result is expected if prominent word-initial vowels are likely to be preceded by a glottal stop gesture. In contrast, word-initial sonorants and their following vowels do not show laryngealized phonation when prominent. Indeed, they are also never preceded by a full glottal stop.

On the other hand, phrasing does not seem to affect the voice quality of initial vowels differently than that of initial sonorants. Instead, all voiced segments show noisier phonation at the onset of higher prosodic domains. This does not mean that phrasing never accounts for the presence of a glottal stop; about 20% of phrase-initial but non-prominent vowels are preceded by full glottal stops. However, in general, onsets of higher prosodic domains yield phonation that is more characteristic of breathiness, at least based on acoustic data. In the following section, I will test whether articulatory measures show similar effects of phrasing and prominence.

## 5.3 Articulatory study

The main goals of this laboratory speech study are to confirm (by means of electroglottograhic contact quotient) the interpretation of the acoustic findings from the corpus study, and to see if they can be generalized to another language (here, Spanish). In Spanish, word-initial glottalization is known to be rarer than in English (Bissiri et al., 2011), but the phrasing effects found in the previous study might generalize to Spanish. Thus, the specific hypotheses tested are:

1. In English, word-initial vowels should show increased contact under prominence. Word-initial sonorants and post-sonorant vowels should not.

2. In Spanish, prominence should result in less contact for vowel-initial words than in English, because glottalization is rarer in Spanish.

3. In both languages, higher phrasal domains should show decreased contact.

I will also obtain simultaneous acoustic measures (of H1*-H2*), which will allow for replication of the corpus study in laboratory speech. If the acoustic results mirror those from the corpus study, then I assume that the articulatory measures here can also be used to interpret the corpus results.

### 5.3.1 Method

#### 5.3.1.1 Stimuli

The target words in both languages consisted of vowel- or sonorant-initial proper nouns with two or three syllables. Each target sound appeared with and without primary stress. Target word-initial vowels were [æ, ə, oʊ], and target word-initial sonorants were [m, n, l, j, w]. Sonorant-initial words were followed by the vowels [oʊ], [ɔ], or [ɪ] in English and by [a], [o], or [au] in Spanish. Stressed syllables were intended to attract phrasal prominence by means of a pitch accent, whereas unstressed syllables were considered non-prominent. In

total, there were 16 English target names and 14 Spanish names. The complete list of target words in both English and Spanish is shown in Table 5.10.

Each target word was placed in four distinct positions that were likely to be pronounced in four distinct phrasal positions: Utterance-initially (after a breath), IP-initially after a high boundary tone (H%), ip-initially after a high phrase accent (H-), and ip-medially. In the three Utterance-medial conditions, the target words always followed a fixed vowel ([ɚ] in English and [a] in Spanish). Additionally, in both languages the number of syllables preceding the target word (when utterance-medial) was held constant, as was the total number of syllables (per stress condition), with the exception of the trisyllabic English names *Yolanda* and *Winona*. Thus, in both stress conditions the target syllable was the seventh syllable in the utterance if it occurred utterance-medially. Utterances with stressed target sounds had a total of 15 syllables, whereas those with unstressed target sounds had 15 or 16 in English and 16 in Spanish. The target syllable, if stressed, never bore the nuclear pitch accent of the phrase. The sentence frames in both English and Spanish are shown in Table 5.11. The expected breaks and preceding tones in MAE-ToBI (Beckman & Ayers Elam, 1997) and Mexican Spanish ToBI (de−la−Mota, Butragueño & Prieto, 2010) are indicated.

### 5.3.1.2 Participants

In total, 24 participants were recruited: 12 native speakers of American English (six female and six male), and 12 native speakers of Mexican Spanish (seven female and five male). All English and Spanish speakers were UCLA students, and were awarded course credit for their participation. All native English speakers spoke only English fluently. The average participant's age was 21 (SD = 2.6) for English speakers and 22 (SD = 2.3) for Spanish speakers. The native Spanish speakers also spoke English, though their levels of proficiency in English varied. Five Spanish speakers were raised in Mexico; the remaining speakers were born in the Los Angeles area. All spoke Spanish on a daily basis, and all claimed to be equally or more comfortable speaking Spanish compared to English. The Spanish-speaking participants spoke the *Distrito Federal* variety of Mexican Spanish. To ensure that none of

Table 5.10: Target words in English and Spanish. The target sounds are underlined.

(a) Target words in English.

|  | Sound | Stressed syllable | Unstressed syllable |
|---|---|---|---|
| Initial vowel | [æ, ə] | _A_nna | _A_nnette |
|  | [i] | _I_gor | _Y_vette |
|  | [oʊ] | _O_din | _O_dette |
| Initial sonorant | [m] | _M_organ | _M_aureen |
|  | [n] | _N_ora | _N_oreen |
|  | [l] | _L_aura | _L_oraine |
|  | [j] | _Y_oko | _Y_olanda |
|  | [w] | _W_innie | _W_inona |

(b) Target words in Spanish.

|  | Sound | Stressed syllable | Unstressed syllable |
|---|---|---|---|
| Initial vowel | [a] | _A_na | _A_nita |
|  | [e] | _E_va | _E_vita |
|  | [o] | _O_lga | _O_limpia |
| Initial sonorant | [m] | _M_aya | _M_arina |
|  | [n] | _N_ana | _N_anita |
|  | [l] | _L_aura | _L_aurita |
|  | [j, ɟ] | _Y_ola | _Y_olanda |

the Spanish speakers spoke English-accented Spanish, the Spanish recordings were labeled by a native speaker who confirmed that all speakers spoke unaccented Mexican Spanish.

### 5.3.1.3   Procedure

The task consisted of recorded read speech in either English or Spanish. Each participant read all the target words in four phrasal conditions, and each sentence was repeated twice. Thus, each native English speaker said 60 sentences twice for a total of 120 repetitions, and each native Spanish speaker read 56 sentences twice, for a total of 112 repetitions. The order of the sentences was randomized, such that no two participants read the sentences in the same order. Participants were instructed to read each sentence aloud as naturally as possible, with no special emphasis on a particular word. The recordings were not ToBI-labeled, but

Table 5.11: Sentence frames in English and Spanish. The location of the target word is marked by '*X*'. See Table 5.10 for target words.

(a) Sentence frames in English.

| Phrasal condition | Preceding tone | Preceding break index | Sentence frame |
|---|---|---|---|
| Utterance-initial | (none: breath) | '5' | X *was sitting on the sofa for the entire day.* |
| IP-initial | H% | 4 | *Was that Alexander?* X *was talking to him today.* |
| ip-initial | H- | 3 | *Teddy, Alexander,* X*'s older sister, and Jim slept.* |
| ip-medial | (L+)H* | 1 | *Alex liked to bother* X*'s older sister on the trip.* |

(b) Sentence frames in Spanish.

| Phrasal condition | Preceding tone | Preceding break index | Sentence frame and translation |
|---|---|---|---|
| Utterance-initial | (none: breath) | '5' | X *estuvo sentada casi todo el día.* 'X was sitting for nearly the whole day.' |
| IP-initial | H% | 4 | *¿Viste a María?* X *no puede encontrarla.* 'Did you see María? X can't find her.' |
| ip-initial | H- | 3 | *Paulina Rivera,* X *Cuarón y yo nos fuimos.* 'Paulina Rivera, X Cuarón, and I left.' |
| ip-medial | L+(>)H* | 1 | *En el parque vi a* X *con su hermano Pedro.* 'I saw X in the park with her brother Pedro.' |

labelers listened to them and noted if the preceding break or tone differed from the intended category (as shown in Table 5.11) in any repetition, in which case that token was excluded. If participants read a sentence with focus on the target word, such that it bore a nuclear pitch accent, or if they read a sentence with too small or large a preceding break, they were asked to repeat the sentence. Any readings with unexpected Utterance-medial breaks or dysfluencies were also excluded, and participants were then asked to repeat the sentence.

The participants were recorded in a sound-attenuated room at UCLA using a Shure SM10A head-mounted microphone in Audacity at a sampling rate of 22,050 Hz. Simultaneous EGG recordings (as a second channel in a stereo audio recording) were collected at the same sampling rate using a two-channel Glottal Enterprises electroglottograph (Model EG2), with a high-pass filter of 20 Hz. The recording lasted approximately 30-45 minutes.

### 5.3.1.4 Labeling and measures

The target sounds were labeled and extracted for acoustic and electroglottographic analyses. In the case of word-initial sonorants, the post-sonorant vowel was also extracted for subsequent analysis, following the same criteria used in the corpus study (and extended to Spanish). For ip-medial sentences with *Ana* in Spanish (in the sentence *En el parque vi a Ana...*), often the first [a] (from the preposition 'a') was difficult to distinguish from the [a] in 'Ana' (indeed, the preposition was often elided). If the labelers heard the two instances of [a] as distinct, the boundary was taken to be the middle of the long [a] sequence. Otherwise, if they heard only one [a], it was attributed to the target word.

The audio waveforms of the extracted segments were then analyzed for acoustic measures using VoiceSauce (Shue et al., 2011), as described above. The electroglottographic waveforms were analyzed for EGG measures using EggWorks, a free EGG analysis program created by Henry Tehrani at UCLA. The subsequent analysis will focus on electroglottographic contact quotient as the articulatory correlate of vocal fold contact, and H1*-H2* (or uncorrected H1-H2 in the case of sonorants) as the acoustic correlate of vocal fold contact. Contact quotient (CQ) was measured using the hybrid method, which defines the point of vocal fold closure as the peak in the derivative of the EGG signal (following Howard (1995)), and uses a 25% peak-to-peak amplitude threshold for detecting the point of vocal fold opening (following Orlikoff (1991)). This hybrid method for measuring contact was used because thresholds at 20% and 25% are found to be best correlated with contact measured via direct imaging of the glottis (Herbst & Ternström, 2006), and because this particular version of EGG contact quotient was found to be most sensitive to changes in voice quality (Kuang, 2011). CQ, H1*-H2*, and H1-H2 were standardized within speaker, and outliers (absolute $Z$-score $> 3$) were removed.

### 5.3.2 Results

#### 5.3.2.1 English results

In the analyses that follow, the statistical difference in mean value for H1*-H2* or H1-H2 is assessed by linear mixed-effects modeling. As in the corpus study, the acoustic or EGG measure is the dependent variable, and prosodic condition*prominence (main effects and interaction) are the fixed effects. The models also had the same random structure as the models in Section 5.2. The results for word-initial vowels are shown in Figure 5.8. CQ (a) and H1*-H2* (b) are plotted as a function of prominence and phrasal domain. For prominent vowels, no difference as a function of prosodic domain is found. For non-prominent vowels, there is a significant main effect of phrasing, whereby higher domains are associated with lower values of CQ (less constriction). The difference in CQ as a function of vowel prominence is significant at all phrasal levels above the ip. The main effect of prominence on CQ (across all phrasal domains) is also significant. Surprisingly, none of the differences in H1*-H2* between prominent vs. non-prominent vowels was significant, as can be seen in Figure 5.8b (cf. results from Section 5.2). $T$-values can be found in Table 5.12.



(a) CQ          (b) H1*-H2*

Figure 5.8: Mean CQ (a) and H1*-H2* (b) for word-initial vowels in English. Prominent vowels are the dashed lines. Error bars indicate 95% confidence intervals.

Table 5.12:  *T*-values for pairwise comparisons of (standardized) CQ and H1*-H2* within and across prominence groups for word-initial vowels in English (see also Figure 5.8). Values below -2 or greater than 2 are considered significant.

(a) Difference in CQ within prominence.

|  | ip-medial vs. ip-initial | ip-initial vs. IP-initial | IP-initial vs. Utt-initial | Main effect-phrasing |
|---|---|---|---|---|
| Prominent | -0.50 | -0.50 | 0.94 | -0.22 |
| Non-prominent | 0.83 | 1.02 | 1.06 | **-2.80*** |

(b) Difference in CQ for prominent vs. non-prominent initial vowels.

| ip-medial | ip-initial | IP-initial | Utt-initial | Main effect of prominence |
|---|---|---|---|---|
| 0.22 | 1.56 | **3.21*** | **2.98*** | **3.97*** |

(c) Difference in H1*-H2* within prominence.

|  | ip-medial vs. ip-initial | ip-initial vs. IP-initial | IP-initial vs. Utt-initial | Main effect-phrasing |
|---|---|---|---|---|
| Prominent | -0.34 | 1.04 | -0.32 | 0.42 |
| Non-prominent | -1.22 | -0.54 | -0.77 | 1.22 |

(d) Difference in H1*-H2* for prominent vs. non-prominent initial vowels.

| ip-medial | ip-initial | IP-initial | Utt-initial | Main effect of prominence |
|---|---|---|---|---|
| 0.14 | -0.12 | -0.57 | -0.72 | -0.33 |

The results for word-initial sonorants are shown in Figure 5.9. Recall that uncorrected H1-H2 is used here because of problems with the formant correction during sonorants. Regardless of prominence, ip-medial sonorants have higher contact and, correspondingly, lower values of H1-H2 than ip-initial sonorants (see Table 5.13). In addition, ip-initial sonorants have higher CQ values than IP-initial ones, though no corresponding change in H1-H2 is found. The main effects of phrasing on CQ and H1-H2 are significant for prominent and (for H1-H2) for non-prominent sonorants. Higher prosodic domains are associated with a decrease in CQ and an increase in H1-H2. No effect of prominence is found on either CQ or H1-H2.

114

(a) CQ           (b) H1-H2 (uncorrected)

Figure 5.9: Mean CQ (a) and H1-H2 (b) for word-initial sonorants in English. Prominent sonorants are the dashed lines. Error bars indicate 95% confidence intervals.

Table 5.13: *T*-values for pairwise comparisons of (standardized) CQ and H1-H2 within and across prominence groups for word-initial sonorants in English (see also Figure 5.9). Values below -2 or greater than 2 are considered significant.

(a) Difference in CQ within prominence.

|  | ip-medial vs. ip-initial | ip-initial vs. IP-initial | IP-initial vs. Utt-initial | Main effect-phrasing |
|---|---|---|---|---|
| Prominent | **2.94***| **4.10*** | 0.51 | **-2.62*** |
| Non-prominent | **2.05*** | **3.53*** | 0.45 | -1.88 |

(b) Difference in CQ for prominent vs. non-prominent sonorants.

| ip-medial | ip-initial | IP-initial | Utt-initial | Main effect of prominence |
|---|---|---|---|---|
| 0.72 | 0.09 | -0.39 | -0.39 | 0.10 |

(c) Difference in (uncorrected) H1-H2 within prominence.

|  | ip-medial vs. ip-initial | ip-initial vs. IP-initial | IP-initial vs. Utt-initial | Main effect-phrasing |
|---|---|---|---|---|
| Prominent | **-2.32*** | -0.08 | 0.91 | **2.85*** |
| Non-prominent | **-2.89*** | -1.28 | 0.23 | **2.89*** |

(d) Difference in (uncorrected) H1-H2 for prominent vs. non-prominent sonorants.

| ip-medial | ip-initial | IP-initial | Utt-initial | Main effect of prominence |
|---|---|---|---|---|
| 0.24 | -0.11 | -0.90 | -1.30 | -0.64 |

(a) CQ  (b) H1*-H2*

Figure 5.10: Mean CQ (a) and H1*-H2* (b) for post-sonorant vowels in English. Prominent vowels are the dashed lines. Error bars indicate 95% confidence intervals.

Table 5.14:   *T*-values for pairwise comparisons of (standardized) CQ and H1*-H2* within and across prominence groups for post-sonorant vowels in English (see also Figure 5.10). Values below -2 or greater than 2 are considered significant.

(a) Difference in CQ within prominence.

|  | ip-medial vs. ip-initial | ip-initial vs. IP-initial | IP-initial vs. Utt-initial | Main effect-phrasing |
|---|---|---|---|---|
| Prominent | **3.04*** | 0.22 | 0.03 | **-2.86*** |
| Non-prominent | **2.00*** | **2.14*** | 0.87 | -1.72 |

(b) Difference in CQ for prominent vs. non-prominent post-sonorant vowels.

| ip-medial | ip-initial | IP-initial | Utt-initial | Main effect of prominence |
|---|---|---|---|---|
| 0.55 | -0.07 | 1.11 | 1.60 | 0.93 |

(c) Difference in H1*-H2* within prominence.

|  | ip-medial vs. ip-initial | ip-initial vs. IP-initial | IP-initial vs. Utt-initial | Main effect-phrasing |
|---|---|---|---|---|
| Prominent | -0.82 | 1.69 | 0.58 | 1.00 |
| Non-prominent | -1.12 | 0.02 | -0.30 | 0.99 |

(d) Difference in H1*-H2* for prominent vs. non-prominent post-sonorant vowels.

| ip-medial | ip-initial | IP-initial | Utt-initial | Main effect of prominence |
|---|---|---|---|---|
| -0.37 | -0.58 | -1.83 | **-2.44*** | -1.67 |

We now turn to the vowels following initial sonorants. Recall that these vowels are meant to provide two kinds of information: (1) whether phrasal effects last throughout the initial syllable, and (2) whether non-initial vowels behave similarly to initial ones, with regards to both prominence and phrasing effects. In terms of phrasing effects, ip-medial post-sonorant vowels have higher CQ contact than ip-initial ones. There is also a main effect of phrasing on CQ for prominent post-sonorant vowels, with a decrease in CQ as a function of increasingly-high phrasal domain. Similarly to what was found in the corpus study, no phrasing effects on H1*-H2* are significant. No effects of prominence on CQ are significant, but significantly lower H1*-H2* values under prominence are found in Utterance-initial position. However, no main effect of prominence is found for either measure. The results support the assumption that phrasal effects will last beyond the word-initial segment; there is a significant drop in CQ as a function of an increasingly-high prosodic domain (but only for prominent vowels). That post-sonorant vowels do not show a main effect of prominence is similar to the result in the corpus study. However, effects of prominence on H1*-H2* are indeed found, though only at the highest phrasal position.

#### 5.3.2.2  Spanish results

The results for word-initial vowels in Spanish are shown in Figure 5.11. CQ (left) and H1*-H2* (right) are plotted as a function of prominence and phrasal domain. For both prominent and non-prominent vowels, higher values of CQ and lower values of H1*-H2* are found ip-medially compared to ip-initial position, and there is a significant corresponding main effect of phrasing on both measures. CQ is higher for non-prominent Utterance-initial vowels than for non-prominent IP-initial vowels, likely indicating increased glottalization of non-prominent initial vowels at the highest prosodic domain. Prominent initial vowels have increased CQ and lower H1*-H2* IP- and Utterance-initially. Prominent vowels have lower values of H1*-H2* also at the ip-initial level. There is an overall main effect of prominence on CQ, with higher values found for prominent vowels than for non-prominent ones. A

corresponding main effect of prominence is found for H1*-H2*, with lower values of the measure under prominence. *T*-values can be found in Table 5.15.
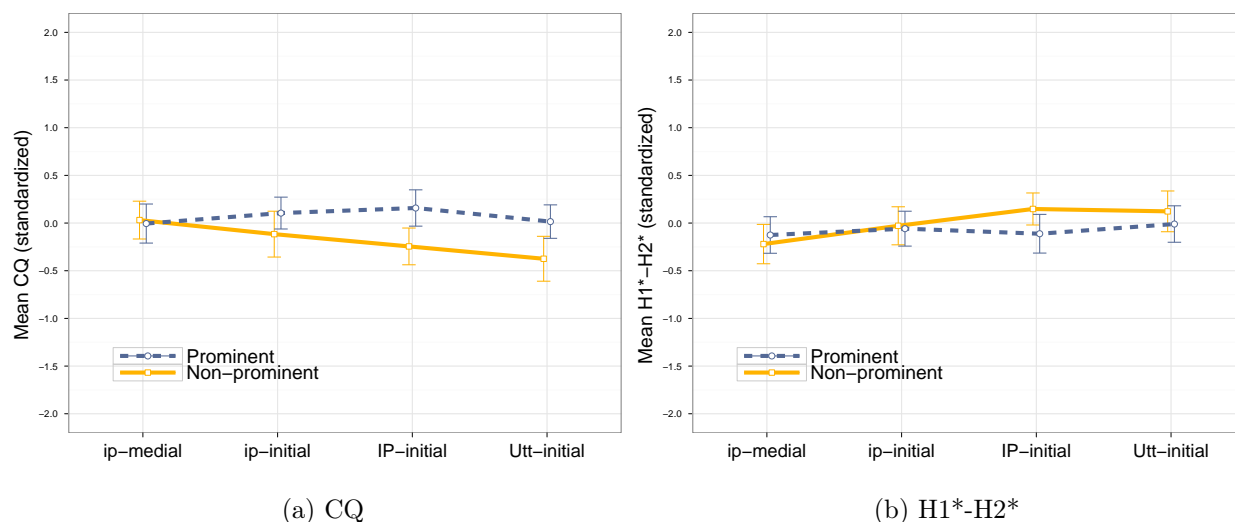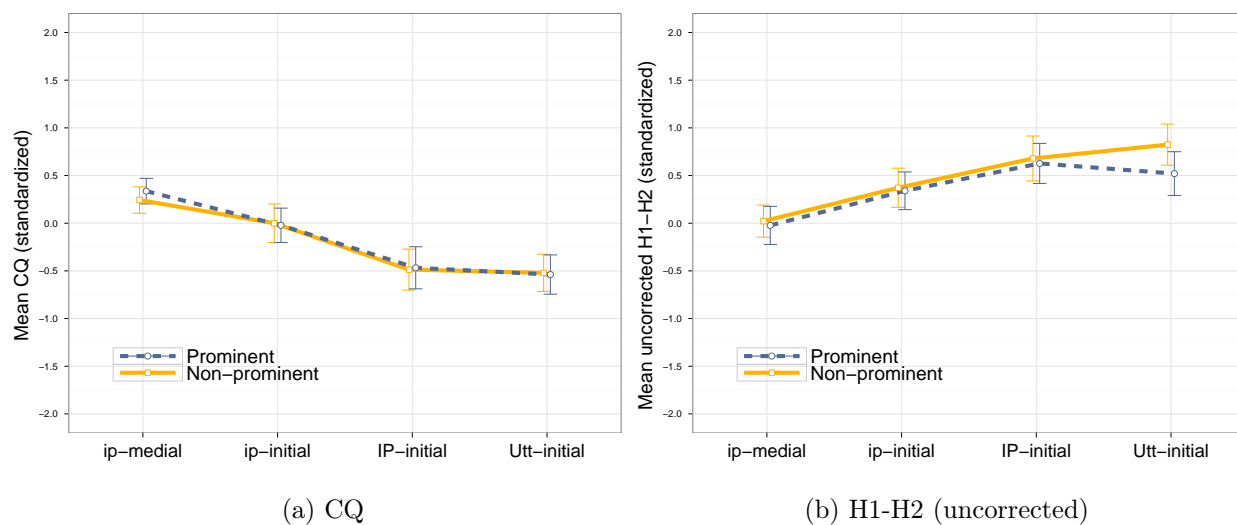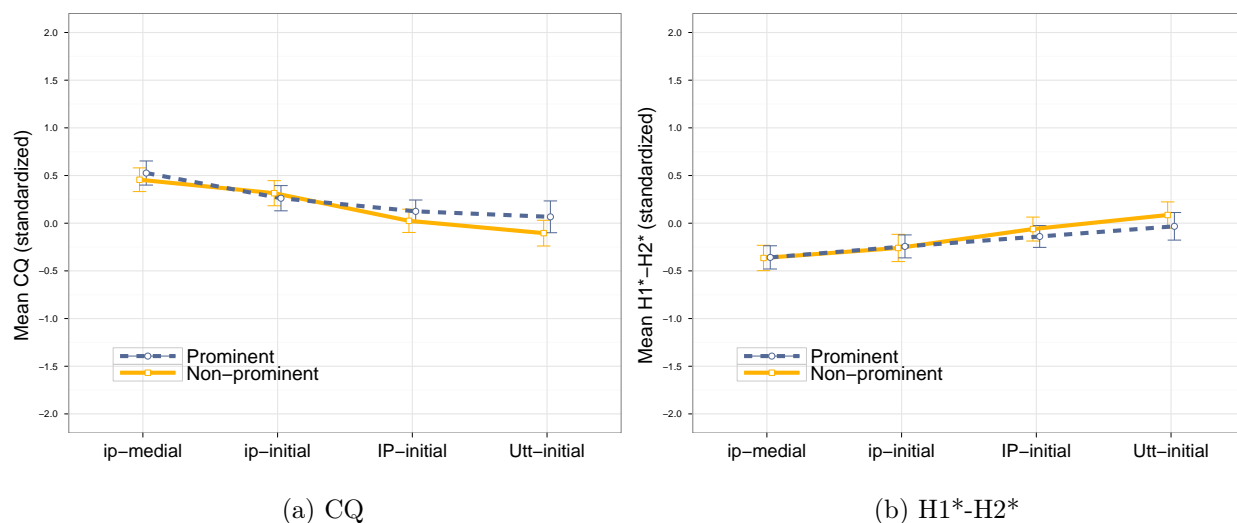


(a) CQ                    (b) H1*-H2*
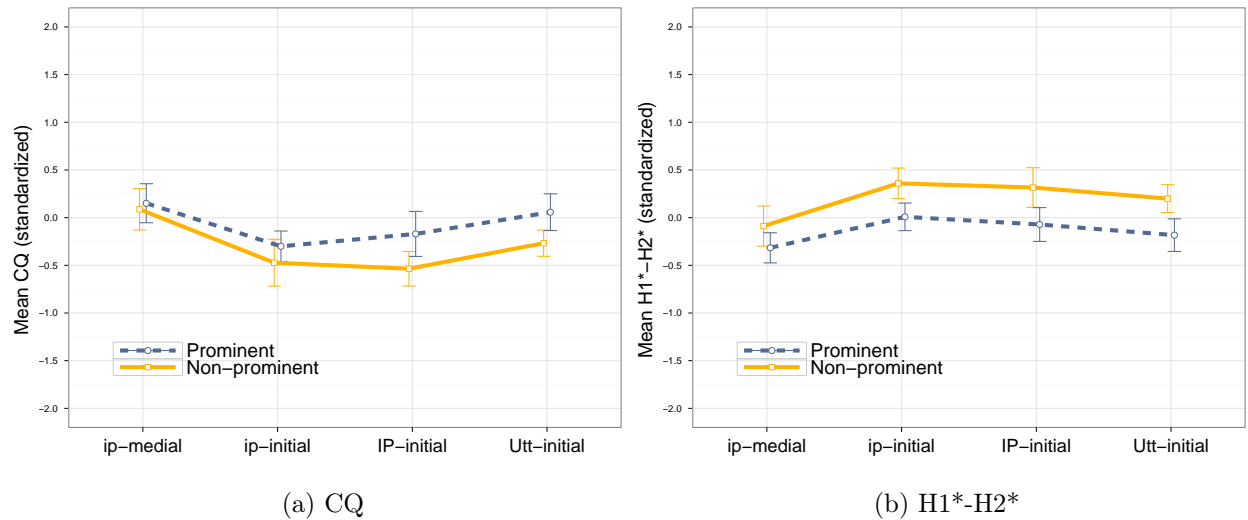
Figure 5.11: Mean CQ (a) and H1*-H2* (b) for word-initial vowels in Spanish. Prominent vowels are the dashed lines. Error bars indicate 95% confidence intervals.

The results for word-initial sonorants in Spanish are shown in Figure 5.12 and Table 5.16. CQ (left) and (uncorrected) H1-H2 (right) are plotted as a function of prominence and phrasal domain. When non-prominent, ip-medial sonorants have higher values of CQ than ip-initial ones, though both prominent and non-prominent ip-initial sonorants have corresponding lower values of H1-H2 than ip-initial ones, which in turn have lower values than IP-initial ones. Non-prominent sonorants also show a significant decrease in H1-H2 at the Utterance-initial level. A significant main effect of phrasing on CQ is found for non-prominent sonorants, with CQ decreasing as a function of higher prosodic domain. The main effect of phrasing on H1-H2 is significant for both prominent and non-prominent sonorants.

The results for post-sonorant vowels in Spanish are shown in Figure 5.13. CQ (left) and H1*-H2* (right) are plotted as a function of prominence and phrasal domain. The only significant effect of phrasing is the decrease in CQ from prominent ip-medial post-sonorant vowels to prominent ip-initial vowels. No significant main effect of phrasing is found. At no prosodic level is there a significant effect of prominence on CQ, and no main effect of

118

Table 5.15: *T*-values for pairwise comparisons of (standardized) CQ and H1\*-H2\* within and across prominence groups for word-initial vowels in Spanish (see also Figure 5.11). Values below -2 or greater than 2 are considered significant.

(a) Difference in CQ within prominence.

|  | ip-medial vs. ip-initial | ip-initial vs. IP-initial | IP-initial vs. Utt-initial | Main effect-phrasing |
|---|---|---|---|---|
| Prominent | **3.44**\* | -1.02 | -1.63 | **-3.79**\* |
| Non-prominent | **4.61**\* | 0.12 | **-2.06** | **-5.03**\* |

(b) Difference in CQ for prominent vs. non-prominent initial vowels.

| ip-medial | ip-initial | IP-initial | Utt-initial | Main effect of prominence |
|---|---|---|---|---|
| 0.37 | 1.55 | **2.53**\* | **2.23**\* | **2.61**\* |

(c) Difference in H1\*-H2\* within prominence.

|  | ip-medial vs. ip-initial | ip-initial vs. IP-initial | IP-initial vs. Utt-initial | Main effect-phrasing |
|---|---|---|---|---|
| Prominent | **-2.56**\* | 1.42 | 0.91 | **2.74**\* |
| Non-prominent | **-4.52**\* | 1.43 | 0.00 | **4.59**\* |

(d) Difference in H1\*-H2\* for prominent vs. non-prominent initial vowels.

| ip-medial | ip-initial | IP-initial | Utt-initial | Main effect of prominence |
|---|---|---|---|---|
| -0.53 | **-2.65**\* | **-2.77**\* | **-2.65**\* | **-2.13**\* |

prominence on CQ is found. However, the results for H1\*-H2\* reveal that ip-medial and IP-initial post-sonorant vowels show lower values of the measure under prominence. There is also a significant main effect of prominence overall, with prominent post-sonorant vowels having lower values of H1\*-H2\* than non-prominent ones. *T*-values can be found in Table 5.17.

(a) CQ



(b) H1-H2 (uncorrected)
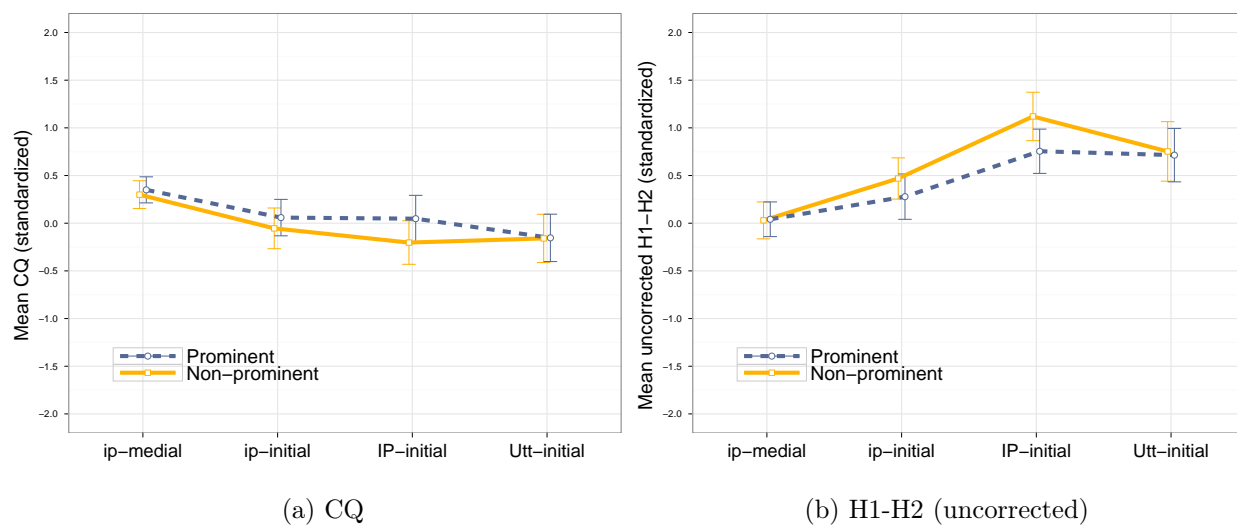
Figure 5.12: Mean CQ (a) and H1-H2 (b) for word-initial sonorants in Spanish. Prominent sonorants are the dashed lines. Error bars indicate 95% confidence intervals.

Table 5.16: *T*-values for pairwise comparisons of (standardized) CQ and H1-H2 within and across prominence groups for word-initial sonorants in Spanish (see also Figure 5.12). Values below -2 or greater than 2 are considered significant.

(a) Difference in CQ within prominence.

|  | ip-medial vs. ip-initial | ip-initial vs. IP-initial | IP-initial vs. Utt-initial | Main effect-phrasing |
|---|---|---|---|---|
| Prominent | 1.55 | 0.39 | 1.52 | -1.20 |
| Non-prominent | **2.32**\* | 1.43 | -0.33 | **-2.35**\* |

(b) Difference in CQ for prominent vs. non-prominent sonorants.

| ip-medial | ip-initial | IP-initial | Utt-initial | Main effect of prominence |
|---|---|---|---|---|
| 0.45 | 1.16 | **2.07**\* | 0.21 | 1.77 |

(c) Difference in (uncorrected) H1-H2 within prominence.

|  | ip-medial vs. ip-initial | ip-initial vs. IP-initial | IP-initial vs. Utt-initial | Main effect-phrasing |
|---|---|---|---|---|
| Prominent | **-2.69**\* | **-2.46**\* | 1.68 | **3.18**\* |
| Non-prominent | **-3.81**\* | **-3.62**\* | **2.55**\* | **4.32**\* |

(d) Difference in (uncorrected) H1-H2 for prominent vs. non-prominent sonorants.

| ip-medial | ip-initial | IP-initial | Utt-initial | Main effect of prominence |
|---|---|---|---|---|
| -0.32 | -0.88 | -1.46 | -0.83 | -1.02 |

(a) CQ

(b) H1*-H2*
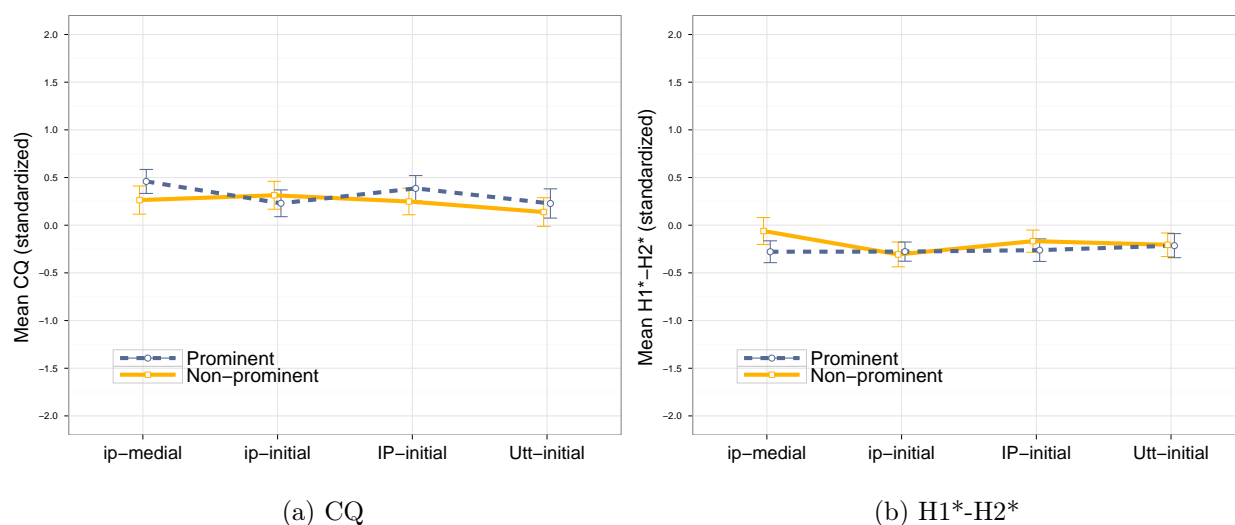
Figure 5.13: Mean CQ (a) and H1*-H2* (b) for post-sonorant vowels in Spanish. Prominent vowels are the dashed lines. Error bars indicate 95% confidence intervals.

Table 5.17: *T*-values for pairwise comparisons of (standardized) CQ and H1*-H2* within and across prominence groups for post-sonorant vowels in Spanish (see also Figure 5.13). Values below -2 or greater than 2 are considered significant.

(a) Difference in CQ within prominence.

|  | ip-medial vs. ip-initial | ip-initial vs. IP-initial | IP-initial vs. Utt-initial | Main effect-phrasing |
|---|---|---|---|---|
| Prominent | **2.54**\* | -1.90 | 1.44 | -1.98 |
| Non-prominent | -0.79 | 0.55 | 1.30 | 1.18 |

(b) Difference in CQ for prominent vs. non-prominent post-sonorant vowels.

| ip-medial | ip-initial | IP-initial | Utt-initial | Main effect of prominence |
|---|---|---|---|---|
| 1.43 | -0.76 | 0.85 | 0.74 | 0.69 |

(c) Difference in H1*-H2* within prominence.

|  | ip-medial vs. ip-initial | ip-initial vs. IP-initial | IP-initial vs. Utt-initial | Main effect-phrasing |
|---|---|---|---|---|
| Prominent | -0.53 | 1.60 | 0.67 | 0.22 |
| Non-prominent | 1.77 | -0.75 | 1.47 | -1.44 |

(d) Difference in H1*-H2* for prominent vs. non-prominent post-sonorant vowels.

| ip-medial | ip-initial | IP-initial | Utt-initial | Main effect of prominence |
|---|---|---|---|---|
| **-2.15**\* | -0.71 | **-2.02**\* | -1.71 | **-2.02**\* |

### 5.3.3 Discussion of articulatory study

Three hypotheses were outlined at the start of this study. First, word-initial vowels (but not initial sonorants) in English and Spanish should show increased contact under prominence. Second, prominence in Spanish should result in a smaller increase in contact for vowel-initial words than in English, because glottalization is rarer in Spanish. Lastly, in both languages, higher phrasal domains should show decreased contact. The results from this study support these hypotheses. Generally, word-initial vowels that are prominent show increased vocal fold contact, though the increase is smaller in Spanish. Phrase-initial voicing in both English and Spanish is characterized by a decrease in contact. In the following section, I discuss the results from this study and the corpus study in Section 2, and implications of both studies for theories of word-initial glottalization.

## 5.4 General discussion

### 5.4.1 Effects of prominence on voice quality

The acoustic effects of prominence on voice quality from the corpus study are largely confirmed in the EGG study. In English, only word-initial vowels show a consistent effect of prominence on voice quality: increased contact. Although the two languages behave similarly, the effect of prominence on word-initial vowel contact and H1*-H2* is greater in degree and more robust across the prosodic hierarchy in English than it is in Spanish, supporting the idea that word-initial glottalization in English is more common than in Spanish (Bissiri et al., 2011). Glottalization of vowel-initial words can be seen as an intentional form of prosodic (specifically, prominence) strengthening. It can be hypothesized that to mark an initial vowel as prominent, speakers deliberately produce stronger voicing.

Alternatively, word-initial glottalization as a form of prominence strengthening could be an unintentional result of a more forceful articulation (Fougeron, 2001, citing also Straka (1963) and Fujimura (1990)). Thus, word-initial glottal stops could be produced because prominent vowels are more forcefully articulated, not because of any linguistically-motivated purpose like contrast enhancement. However, this line of analysis is problematic, because we should also expect prominent sonorants and even non-prominent voiced segments to show increased contact. Thus, under this account it is difficult to explain the different effects found for vowels vs. sonorants and for prominence vs. initial position.

In sum, prominence strengthening in English and Spanish is realized via increased glottal constriction for word-initial vowels. This is consistent either with glottal stop insertion, or with gradient increase in glottal constriction during voicing, which in extreme cases would result in [ʔ]. The latter possibility still requires positing an inserted glottal adduction feature or gesture (e.g., [+ glottalization]), because glottalization is specific to word-initial vowels that are prominent.

### 5.4.2   Effects of phrasing on voice quality

The results of the two studies indicate that vowel and sonorant voicing at the onsets of higher prosodic domains shows a decrease in contact quotient and a corresponding increase in H1*-H2* or H1-H2. Why would voicing show less contact at higher phrasal domains? In this section, I propose several possibilities for why phrase-initial abduction occurs.

Initial strengthening is a possible explanation for phrase-initial abduction. However, strengthening that results in a more forceful articulation (Fougeron, 2001) cannot trigger the phrasal effects found here, because a more forceful articulation is expected to result in laryngealization, rather than in breathy-like phonation.

On the other hand, during initial strengthening the *contrast* between the target sound and its neighbors could be strengthened ('syntagmatic enhancement'; see discussion by Fougeron & Keating (1997); Hsu & Jun (1998); Fougeron (2001), and Cho (2005), among others).

Syntagmatic enhancement does not necessarily result in a more forceful or prototypical articulation. For example, the decrease in duration and amount of nasal flow for phrase-initial nasals in French (Fougeron, 2001) can be viewed as a form of syntagmatic enhancement: phrase-initial nasals are more consonantal and thus less similar to the following vowel, even though they become less 'nasal.' Likewise, because voice quality at phrasal onsets shows a decrease in contact, word-initial vowels and sonorants can be viewed as becoming more consonantal (breathier, more [h]-like). This interpretation however runs into problems when we compare word-initial sonorants to post-sonorant vowels. In English and (to a lesser degree) in Spanish, post-sonorant vowels also show decreased vocal fold contact at higher prosodic domains (see Figures 5.10 and 5.13). If the purpose of decreasing vocal fold contact phrase-initially is to enhance the contrast between the initial segment and what follows, then we cannot explain why the entire initial syllable – not just the initial segment – shows the effect.

Another possible explanation for the decrease in contact phrase-initially is that it is due to respiratory constraints on voicing initiation. Indeed, Slifka (2000) provides evidence that glottal area increases at voicing initiation. This might be deliberate on the part of the speaker, because an increase in glottal opening would reduce phonation onset time by increasing transglottal airflow. But the increase in glottal area might also be unintentional, in that during inspiration the vocal folds open widely to allow air to pass to the lungs. Thus, possible respiratory reasons for decreased vocal fold contact are (1) that speakers increase glottal opening to increase transglottal flow and decrease phonation onset time, or (2) that the glottis is more open Utterance-initially due to the preceding intake of breath. However, neither explanation based on respiratory constraints is entirely satisfactory, because all phrasal onsets show a decrease in contact. Indeed, the most consistent effect of phrasing on voice quality is the difference between ip-medial (word-initial) and ip-initial domains, which are both Utterance-medial. Onsets of Utterance-medial phrases should not be influenced by respiration, at least not for the reasons mentioned above. No pauses occurred at intermediate

124

phrase boundaries, and the ip-initial vowels and sonorants were preceded by a voiced segment, meaning that voicing did not have to 'restart' at ip onsets.

It is also possible that the Utterance-initial effects are generalized and phonologized down the prosodic hierarchy. This would then be similar to coda obstruent devoicing: the aerodynamic/laryngeal conditions favoring coda-devoicing are only found Utterance-finally (Westbury & Keating, 1986), yet the process often generalizes to all word-final or even syllable-final positions in some languages (see Myers (2012) for a review of languages). Generalization of decreased contact for Utterance-initial voicing to all phrasal onsets could provide listeners with a consistent perceptual cue to phrase onsets, though it is currently unclear whether listeners attend to such changes in voice quality.

Unrelated to respiratory constraints, a final possible reason for phrase-initial abduction is pitch reset. Pitch reset is the change in the slope of F0 declination (Ladd, 1984, 2008). The domain of pitch reset is often assumed to be at the level of the ip, though there is evidence for increased pitch reset at higher domain onsets (O'Shaughnessy & Allen, 1983; Ladd, 1988, 2008). Crucially, all phrase onsets are accompanied by pitch reset. I propose that pitch reset could trigger increased vocal fold abduction if it involves muscular relaxation of the thyroarytenoid (TA) and/or cricoarytenoid (CT). Indeed, there is evidence of TA (vocalis) and CT relaxation when F0 falls (Hirano, Ohala & Vennard, 1969), which we would expect given increased CT and TA activation at higher pitch levels (Hirose, 1997; Kreiman & Sidtis, 2011)). Unexpectedly though, Hirano et al. (1969) also found brief relaxation of the vocalis and CT before a sharp rise in F0. CT and TA relaxation results in a decrease in vocal fold adduction, as Mendelsohn & Zhang (2011) and Zhang (2011) show. This would imply that there should be small vocal fold abduction at phrase onsets – not large enough to render phrasal onsets voiceless, but large enough to add breathiness (reminiscent of phrasally-modulated segmental spread glottis (Pierrehumbert & Talkin, 1992; Jun, Beckman & Lee, 1998)). Thus, the rapid change in F0 triggered by pitch reset at phrase boundaries could be responsible for increased vocal fold abduction at all phrasal domains.

In sum, the most likely explanations for phrase-initial effects on voice quality are respiratory constraints on Utterance-Initial phonation (if then phonologized down the prosodic hierarchy), and/or pitch reset effects on vocal fold adduction. Because in this paper I focused on glottalization rather than phrase-initial voice quality, further work is needed to determine the exact mechanism responsible for the consistent effect found here. But whatever the reason, it is clear that phrase-initial (abducted) voice quality is a distinct phenomenon from word-initial glottalization (adduction).

### 5.4.2.1   Non-prominent but phrase-initial glottal stops

It should nonetheless be noted that, in the corpus study, a non-negligible number (roughly 20%) of *non-prominent* initial vowels are preceded by [?] phrase-initially (see Figure 5.3). This is surprising, because there is no effect of phrasing on the articulatory and acoustic measures for word-initial vowels. Why are glottal stops found in this context, but not more laryngealized voice quality?

It is likely that glottal stops (both full and incomplete) can optionally precede phrase-initial *non-prominent* vowels, much as they seem to do for prominent initial vowels. But if a phrase-initial vowel is not glottalized, its default voice quality would be breathier. This would result in some phrase-initial vowels having breathier voice quality (the ones with no glottalization) and some with more laryngealized voice quality (those with incomplete glottal stops). The end result would be no main effect of phrasing on the voice quality of word-initial vowels, as found in the corpus study. If this is true, then we can conclude that phrase-initial vowels may sometimes undergo glottalization (likely due to strengthening), but if they do not, they will 'succumb' to phrase-initial abduction like other voiced sounds. There is some evidence for optional phrase-initial glottalization for non-prominent initial vowels from the bimodal spread of data in BU Radio News Corpus, shown in Figure 5.14. Non-prominent phrase-initial vowels seem to fall under two groups: one with overall lower values of H1*-H2* (corresponding to those that are glottalized), and another with overall higher values of H1*-H2* (corresponding to those that are not glottalized).

126
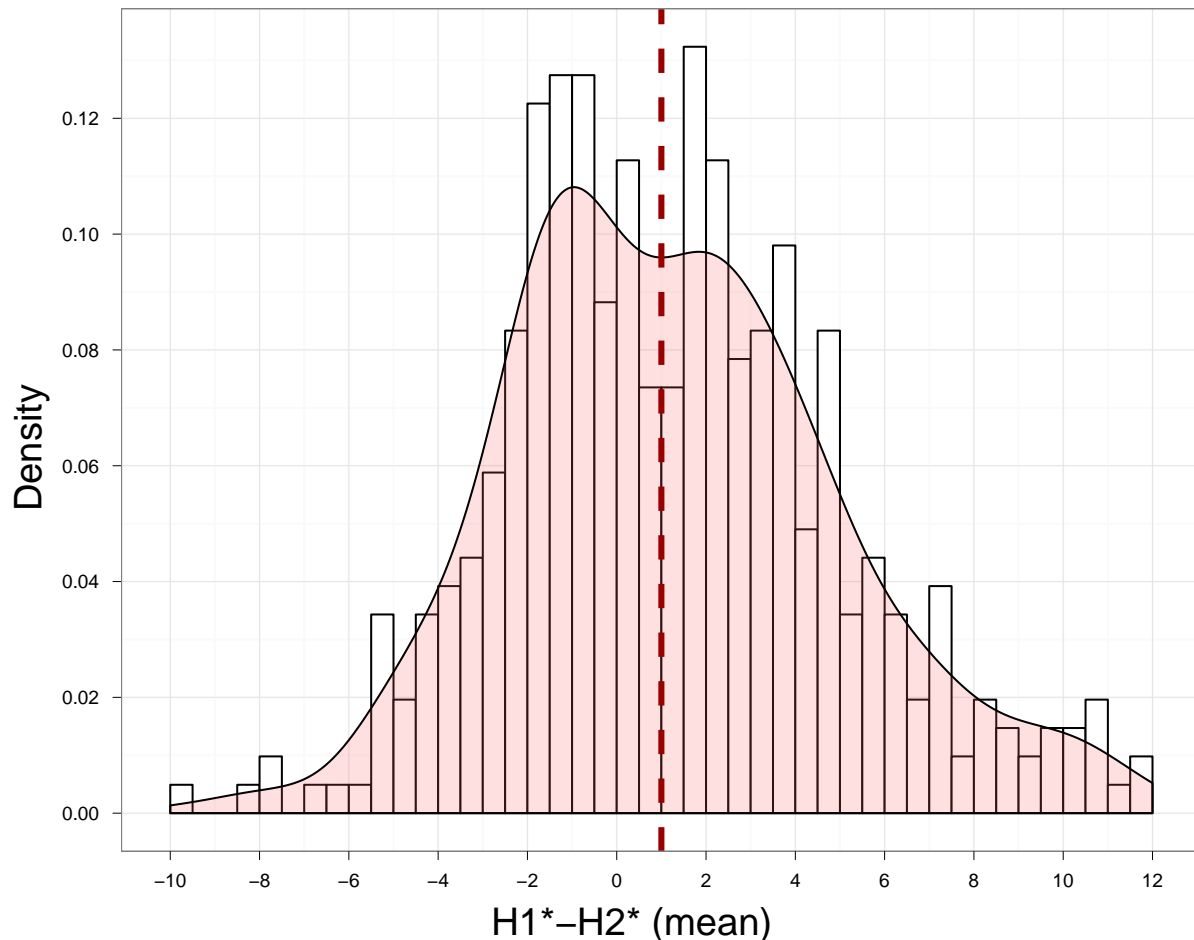
Figure 5.14: Histogram and kernel density estimation of H1*-H2* for non-prominent phrase-initial vowels in the BU Radio News Corpus. Note that the spread appears bimodal, with the approximate boundary between the two subgroups marked with a dotted line.

Another possibility is that incomplete glottal stops do not occur in non-prominent phrase-initial positions, but full glottal stops do. This would explain why non-prominent initial vowels in the EGG study showed a decrease in contact quotient (cf. Figures 5.8a For English and 5.11a for Spanish). If this is true, then full glottal stops in phrase-initial position would serve a different function than those found under prominence. In Section 5.2.2.1, I hypothesized that the increase in [ʔ] occurrence after periods of voicing irregularity might be derived from a form of voicing re-initiation. During voicing irregularity such as phrase-final creak, the vocal folds vibrate unpredictably, with long open and closed periods (Slifka, 2006). Thus, it is possible that full glottal stops are used by some speakers to restart phonation

after the end of a phrase, even if the preceding phrase does not end in strong creak. The use of [ʔ] (rather than [ʔ̰]) would allow for a rapid buildup in subglottal pressure and acoustic energy. Phrase-initial [ʔ]s also make good cues to phrase boundaries, because the silence and burst of the full glottal stop preceding the vowel would provide an auditory boost to listeners (Delgutte, 1980, 1982; Delgutte & Kiang, 1984). On the other hand, an auditory a boost would be much weaker if the vowel were simply laryngealized and not preceding by [ʔ].

### 5.4.3    On the cross-linguistic prevalence of word-initial glottalization

In light of the findings on where glottal stops occur and the dissimilarity between glottal stops and phrase-initial voice quality, it is likely that glottal stops derive from an inserted glottal constriction gesture on word-initial vowels that are prominent, especially when they occur in phrase-initial position. Full [ʔ] occurs before prominent initial vowels more often phrase-initially than phrase-medially because of initial strengthening of the glottal adduction gesture. Thus, unlike previous researchers who viewed word-initial glottalization as a form of prosodic strengthening more generally (Pierrehumbert & Talkin, 1992; Dilley et al., 1996; Borroff, 2007), I claim here that the phenomenon is more accurately viewed specifically as a form of prominence strengthening. Furthermore, there is no articulatory or acoustic evidence that a glottal stop gesture is always present before word-initial vowels in English (a hypothesis discussed in Pierrehumbert & Talkin (1992); Dilley et al. (1996), and Borroff (2007)). As mentioned earlier, we cannot disprove this entirely, but given that glottal stops are not common phrase-initially – despite the fact that phrase-initial position strengthens other segments – this hypothesis is counterintuitive; we do not expect a segment to lenite in strong phrasal positions.

Let us now return to the main research question: why then is word-initial glottalization so common across languages? It cannot be because phrase-initial voicing is in general tenser, because our results show that the opposite is true. Moreover, if phrase-initial vocal fold abduction is due to respiratory and/or muscular constraints, such abduction is likely to

be universal, and would counter any trend to glottalize phrase-initially. But put the other way round, glottal stops can counteract the effects of phrase-initial vocal fold abduction, a countering which is useful for word-initial vowels. Prominent Utterance-initial vowels must convey prominence despite their low and rapidly-changing subglottal pressure and their increased glottal area. Prominent phrase-initial (but Utterance-medial) vowels must convey prominence despite pitch reset, which likely causes a brief period of glottal abduction. Further, breathy voice quality can be detrimental to pitch recoverability (Silverman, 1995, 2003), and prominence is usually marked by tones in languages of the world (Jun, 2005). The purpose of prominence is to convey salient information (Pierrehumbert & Hirschberg, 1990; Ladd, 2008), so to indicate prominence on phrase-initial vowels, the speaker must counteract the effects of phrasing on voice quality. A glottal stop ensures that the initial vowel will be produced with a voice quality that is more conducive to conveying prominence.

Further support for [ʔ] (but not any form of glottalization) comes from findings on auditory adaptation: a glottal stop before prominent vowels would ensure that there is a period of silence and a stop burst before the prominent vowel, which should increase listeners' auditory sensitivity (Delgutte & Kiang, 1984). Under this account though, the default marker of word-initial vowel prominence would be a full glottal stop [ʔ], not glottalization more generally, because a laryngealized vowel would not provide the same auditory boost as a full stop. This in turn implies that a prominent vowel without a full [ʔ] is in fact lenited. Again, this is counterintuitive if we assume that prominent segments are typically produced with greater articulatory strength. At any rate, under both accounts word-initial glottalization can be used as a reaction to phrase-initial voice quality. Initial sonorants in prominent syllables do not need to undergo glottalization, either because prominence can be conveyed on the following vowel, where the influence of phrase-initial position on the voice quality is weaker, or because the rapid decline in energy during sonorants would result in a similar auditory effect as [ʔ].

Why then would word-initial glottalization rates vary across languages? As noted earlier, there are languages in which word-initial glottalization almost always occurs (e.g., Czech;

129

Bissiri & Volín (2010)); others, like English or German, which show less frequent occurrence; and Spanish or French, where it is thought to be even rarer (Fougeron, 2001; Borroff, 2007; Bissiri et al., 2011). Languages with very frequent word-initial glottalization are likely to be stress-initial (e.g., Czech). But even if they do not attract prominence on word-initial vowels, it is possible that prominence-induced glottalization has been generalized to all word-initial vowels, even those that are not prominent. As for cases of infrequent word-initial glottalization, I hypothesize that these languages must cue prominence by other means, e.g. through intonation. For example, in Mexican Spanish the most common pitch accent is L+>H* (with a delayed high tone; de−la−Mota et al. (2010)), meaning that the pitch maximum is reached after the stressed syllable. For vowel-initial words (e.g. 'Ana'), this means that the pitch maximum occurs on the second syllable, where the effects of phrase-initial abduction are reduced. Thus, in Spanish cues to prominence extend beyond the stressed syllable. Moreover, post-lexical prominence can be marked by edge tones (e.g., in Korean, Mongolian, and unaccented Japanese), instead of or in addition to marking the head of the prominent word by means of local changes in amplitude, duration, and pitch (Jun, 2005). If in a given language prominence is marked by edge tones and not by additional suprasegmental features, one could expect that word-initial glottalization would be rare, because edge-marking prominence on phrase-initial vowels would be adequately conveyed through tones. I leave investigation of the typology of word-initial glottalization to further research.

## 5.5   Conclusion

In this chapter, I presented two studies on word-initial glottalization. In the first study, I used logistic mixed-effects regression modeling to predict the occurrence of word-initial full glottal stops ([ʔ]) in an English corpus. The results indicated that prominence and phrasing are the most important factors in predicting full glottal stop occurrence. Moreover, prominent word-initial vowels that were not preceded by [ʔ] showed a decrease in H1*-H2*, an

acoustic correlate of glottal constriction. Surprisingly, non-prominent phrase-initial vowels did not regularly show signs of glottal constriction. These findings were then confirmed using electroglottographic contact quotient, and extended to Spanish. Based on the results, I proposed and motivated a theory of word-initial glottalization where prominence is the driving force in determining whether word-initial vowels show glottalization, and where higher phrasal domains are responsible for the strength of the glottal constriction gesture. I assume that most word-initial glottal stops are derived from strengthening of an epenthetic glottal constriction gesture/feature of word-initial prominent vowels. Most prominent initial vowels show laryngealization, which is taken here to be the acoustic evidence for glottal constriction. When these prominent initial vowels are also phrase-initial, the glottal constriction is more likely to be realized as a full stop, due to phrase-initial strengthening. Typological differences in the way prominence is cued can account for variable rates of glottalization across languages.

# CHAPTER 6

# General discussion

## 6.1 Summary of results and thoughts for future research

The three studies in this dissertation sought to answer the following questions about glottal stops in English:

1. How are glottal stops produced?

2. Do listeners perceive glottal stops differently than creaky voice?

3. When do glottal stops occur word-initially, and why?

### 6.1.1 How are glottal stops produced?

The results from Chapter 3 indicate that glottal stops are produced by the absence of voicing, which is always achieved via complete or partial closure of the glottis by the vocal folds, and sometimes via additional constriction of the ventricular folds. Ventricular incursion too can be complete or partial. Thus, glottal stops are truly glottal sounds, though the degree of glottal and supraglottal constrictions varies by token, speaker, syllable position, and phrasal position. This variability appears in the present work to be mostly random. The one exception is the effect of syllable position: glottal stops tend to be more strongly produced (more glottal and ventricular constriction) as codas than as onsets. As discussed in Chapter 3, this tendency can be due to the fact that the participants in this study were English speakers, and thus were probably more accustomed to producing [ʔ] in word-initial position than in codas. Nevertheless, a similar discrepancy between onset and coda 'glottal stops' has

been found in Thai, where initial vowels are often thought to be preceded by glottal stops (and are represented as such orthographically), yet there is little laryngoscopic evidence for them (Harris, 2001). So, it is also possible that glottal stops require additional effort to be produced in coda position, likely because they require cessation of voicing. Voicing can begin after [ʔ] as long as the requisite myoelastic and aerodynamic conditions are met (see Section 2.2.3), but voicing cessation with [ʔ] is more complicated: the vocal folds must stiffen and constrict sufficiently to avoid abducting from the increasing subglottal pressure. However, they cannot do so too early, or else the production of the preceding sound (the vowel) will be compromised. Thus, voicing cessation into [ʔ] must occur fairly rapidly, possibly requiring additional articulatory strength – hence supraglottal articulation.

Modeling studies can help clarify under what conditions glottal stops will be produced with full vs. incomplete glottal closure, and with/without ventricular incursion. Fortunately, some of this work is currently underway (Moisik & Esling, 2012), and could be enhanced by the investigation of the effects of subglottal pressure, as well as the presence vs. absence of pre-[ʔ] voicing, on the production of glottal stops. In addition, further imaging work with glottal stops produced in more 'casual' speech can provide useful information on their variability in production. However, imaging with more 'casual' speech would require trans-nasal laryngoscopy (so that real words may be pronounced relatively naturally), in addition to the high sampling rate like the 10,000 fps used in the imaging study in Chapter 3. This is currently a technical challenge, but advances in high-speed cameras could make this kind of experimentation possible in the near future.

However, even small datasets like the one in Chapter 3 provide additional useful information, including changes in voicing leading into and out of the glottal stops, which can help clarify the results from the imaging study. For example, one could ask how voicing is altered around glottal stops (viz. by changes in open quotient, speed of closure, presence and size of posterior gaps), and whether certain changes are associated with certain types of glottal stops: those produced with full vs. partial closure of the glottis, those produced with or without ventricular incursion. This in turn could help solve the apparent randomness

with regard to the variability in glottal stop production. One could expect faster and larger changes to voicing around glottal stops that are more 'strongly' articulated, i.e., those with full glottal closure and with ventricular incursion.

In addition, ventricular incursion appears to be used to reinforce glottal closure for the cases of [ʔ] in Chapter 3. Therefore, it might not be strongly correlated with more rapid and extreme changes to voicing around [ʔ], varying instead as a function of subglottal pressure as well as vocal fold stiffness/tension. Voicing too varies as a function of the subglottal pressure, but the quasi-independent effects of pressure vs. voicing changes can be assessed by investigating ventricular incursion rates for glottal stops neighboring vowels with different types of voice quality and loudness levels. Compared with modal voice, breathy voice is characterized by laxer vocal folds, whereas laryngealization is usually produced with stiffer vocal folds (Laver, 1980; Gordon & Ladefoged, 2001). But louder voicing usually results from higher subglottal pressure (Baken & Orlikoff, 2000, table 8-10 (pp. 329–330), and references therein), and different phonation types can be produced at different intensities. Thus, in future work I can test whether ventricular incursion during [ʔ] is more common when voicing around [ʔ] is laryngealized (due to the more 'forceful' vocal fold vibrations), when it is louder, regardless of voice quality (because glottal stop reinforcement is needed due to an increase in pressure), or when voicing is both loud and laryngealized.

### 6.1.2 Do listeners perceive glottal stops differently than creaky voice?

The results from Chapter 4 show that glottal stops (as allophones of /t/) are generally perceived distinctly from creaky voice in English. That is, most instances of glottal stops are not mistaken for creak, and most instances of creak are not mistaken for glottal stops.

Under certain circumstances, however, glottal stops may be misperceived as creak: when they are short in *atlas*-type words (where there is too little closure duration to be a potential cue to listeners), and when they occur post-nasally and word-finally in *dent*-type words (when glottalization cues are strongest during a nasal rather than a vowel). Creak too can

be mistaken for glottal stops, but usually only in *den*-type words; that is, both creak and glottal stops are harder to identify word-finally after nasal codas. But because word-final and post-nasal positions are confounded in this study, it may seem unclear precisely what renders glottal stops and creak mutually confusable in *dent-den* pairs. I argue that the confusability is due to the fact that glottalization is post-nasal (rather than word-final), because previous work has shown that glottalization improves English listeners' identification of word pairs like *bait-bay*, where the word-final /t/ is glottalized (Garellek, 2011).

These findings imply that listeners make use of cues that can distinguish glottal stops and creak. The acoustic analysis in Section 5.2.2.2 in Chapter 5 reveals that voicing around (full) glottal stops is characterized by changes to the harmonic and inharmonic components of the signal, and some of these same characteristics are similar to the voice quality during phrase-final creak (cf. results in Section 4.2.1.1 in Chapter 4). Thus, listeners cannot rely solely on the cues shared by creaky voice and glottalization if they are usually good at perceiving the two distinctly. Which cues then might they be attending to? Both voiced glottalization and creaky voice are characterized by dips in F0 and energy, as well as an increase in noise. If a glottal stop is produced as full [ʔ], then stop closure is a potential cue; indeed, listeners in the study presented in Chapter 4 identified stimuli with [ʔ] more accurately than those with shortened [ʔ̆].

In addition to stop closure duration for glottal stops, it is possible that listeners rely on other cues to distinguish short glottal stops from creaky voice produced during phrase-final creak. Creaky vowels and nasals before [ʔ] were noisier but had greater intensity than those that were not followed by a glottal stop (but still creaky). Thus, relative noise and intensity might be the key cues to glottal stops in creak. The extent to which listeners rely on these measures (as well as other potential cues like F0 and closure duration) can be assessed in future work using resynthesized tokens, for example in a Garner classification experiment (Garner & Felfoldy, 1970). Such an experiment could help determine which specific cues listeners rely on, and how independent or integrated they are.

The findings from Chapter 4 would be further enhanced by better understanding how phrase-final creak and voiced glottalization ([ʔ̬]) differ acoustically. To do so, in future work I can test (for a large number of speakers) the production of word pairs like the stimuli in the *button-bun, atlas-Alice,* and *dent-den* groups, in both normal and creaky phonation. These target words could be placed in carrier sentences, in phrase-medial and – to investigate the effects of creak – phrase-final positions.

### 6.1.3   When do word-initial glottal stops occur, and why?

In Chapter 5 I focused on the distribution of word-initial glottal stops, which commonly occur before initial vowels in many of the world's languages. The results confirm those from previous studies demonstrating that prosodic factors play an important role in English word-initial glottalization (Pierrehumbert & Talkin, 1992; Dilley et al., 1996). Additionally, I found that prominent and phrase-initial positions were the most important factors in predicting the presence of a full glottal stop. I also used quantitative measures of voice quality to show that word-initial vowels are more laryngealized under prominence even when not preceded by [ʔ], which I take to be evidence for incomplete [ʔ̬]. The quantitative analysis was replicated using electroglottography, and extended to Spanish, which shows similar results to English. Thus, prominence induces glottalization on word-initial vowels (either as full [ʔ] or as laryngealization) in both languages. Phrase-initially, however, English and Spanish are generally characterized by a decrease in vocal fold contact. From these results, I propose that glottalization is largely due to prominence (though likely optional at all phrasal onsets), but the strength of glottalization depends on phrasal position, such that full [ʔ] is more common at phrasal onsets compared with phrase-medial positions:

1. Glottalization: $V_{[+prominent]} \rightarrow [+glottalized] \; / \; \# \; \rule{3cm}{0.4pt}$

2. Glottalization strengthening: $[+glottalized] \rightarrow [ʔ] \; / \; [_{Phrase} \; \rule{3cm}{0.4pt}$

Why is glottalization a marker of prominence for word-initial vowels? Prominent Utterance-initial vowels must convey prominence despite their low and rapidly-changing subglottal

pressure and (based on the results in Chapter 5) their increased glottal area. In Utterance-medial position, prominent phrase-initial vowels must convey prominence despite pitch reset, which likely causes a brief period of glottal abduction. It is known that breathy voice quality (such as is found phrase-initially) can be detrimental to pitch recoverability (Silverman, 1995, 2003), and prominence is usually marked by tones in languages of the world (Jun, 2005). Vowel quality may also be harder to perceive in breathy voice quality, because of the decrease in high-frequency energy (Klatt & Klatt, 1990; Hanson et al., 2001). Given that prominence is used to convey salient information (Pierrehumbert & Hirschberg, 1990; Ladd, 2008), speakers might glottalize initial vowels to counteract the effects of phrasing on voice quality. A glottal stop thus ensures that the initial vowel will be produced with a voice quality that is more conducive to conveying prominence. Glottalization was not found for initial sonorants, but initial sonorants in prominent syllables do not need glottalization, because prominence can be conveyed on the following vowel, where the influence of phrase-initial position on the voice quality is probably weaker. (The experiment in Section 5.3 showed fewer significant effects of phrasing for post-sonorant vowels compared with initial sonorants). Furthermore, the lack of glottalization on initial sonorants argues against general phrase-initial strengthening of voice quality. When all these results are taken together, it is clear that prominence strengthening of voice quality is unique to word-initial vowels, and therefore it cannot be a by-product of more forceful voicing. This lends additional support to the notion that glottalization may be used to enhance salience.

If word-initial glottalization is largely due to prominence, then I expect glottalization rates to differ for languages that cue prominence differently or more weakly than English and Spanish. For example, a language that marks prominence largely by means of edge tones (e.g., Korean, Mongolian, unaccented Japanese), should have rare word-initial glottalization, because edge-marking prominence on phrase-initial vowels would be adequately conveyed through tones. On the other hand, languages with fixed word-initial prominence are expected to have higher rates of glottalization (e.g. Czech: Bissiri & Volín (2010)). But even languages with initial prominence might have infrequent glottalization, if prominence is weakly cued in

the language (e.g. Bengali: Khan (2008)). Thus, I expect prominence-driven glottalization to depend on (1) where prominence is cued in a word/phrase (initial vs. non-initial), (2) how it is cued (by marking edges vs. heads of constituents), and (3) how much prominence 'matters' in the language; i.e., is it generally strongly vs. weakly cued? In future work, it would therefore be interesting to replicate the articulatory study in Section 5.3 of Chapter 5 for theoretically-interesting languages like Korean (no head-driven prominence), as well as Bengali vs. Czech (both stress-initial, but weakly-cued stress in Bengali).

It is clear, however, that prominence cannot account for all cases of [ʔ], and so in Chapter 5 I proposed that glottalization may be optional phrase-initially, and that such instances of [ʔ] phrase-initially and/or after creak may be used to reinitiate phonation under conditions of low subglottal pressure and irregular voicing. In future research, this hypothesis can be tested by noting the presence of phrase-initial glottalization for word-initial vowels produced at different levels of subglottal pressure, e.g. in an experiment with both EGG and oral pressure, where oral pressure indexes subglottal pressure.

In general, acoustic studies of glottalization would be improved by coding for full [ʔ] vs. incomplete [ʔ̞] occurrence, and analyzing both subsets separately. However, it is usually not sufficient to base the ([ʔ̞] vs. ∅) distinction solely on visual inspection of the acoustic signal: e.g. 'if silence before initial vowel → [ʔ]; if irregular pitch periods on initial vowel → [ʔ̞]; else, no glottalization.' This is not sufficient because (1) some instances of glottalization may be undetectable by the naked eye, and (2) because [ʔ̞] resembles voicing irregularity that may result from other phenomena (and at the very least, researchers should state why they believe other sources of irregular pitch periods are not considered confounds). Ideally, studies would combine visual inspection with quantitative measures of voice quality, such that a more gradient analysis can be achieved. Future work on automatic detection of glottal stops and glottalization should also rely on both time-varying (e.g. Seid et al. (2012)) and frequency-varying features in the acoustic signal.

### 6.1.4 Phonological research pertaining to glottal stops and glottalization

Although this dissertation dealt primarily with the phonetic characteristics of glottal stops, the results also have implications for phonological research. A common issue with glottal stops in phonological analyses is whether they should be posited in underlying representations – and if not, when to know if [ʔ] is epenthetic (Lombardi, 2002). Given that glottal stops are highly variable in their articulation, I believe that positing a glottal stop phoneme or allophone should first be based on phonological (e.g., from synchronic alternations), rather than phonetic, evidence. Possible phonetic 'evidence' for glottal stops – irregular voicing and/or a plosive burst preceding a word-initial vowel – is not sufficient evidence for the glottal stop's being part of a language's sound system unless other potential causes of such features (e.g., phrasal position, prominence, and other post-lexical phenomena – see Hyman (1988)) are controlled for. Here are a few examples which outline the problem:

- There is phonetic evidence for a glottal stop phoneme or allophone word-initially...

    - Especially if the evidence is derived from recordings of words in isolation, the phonetic signs of glottalization can be due to phrase-initial position.

    - If the initial vowel bears some form of prominence (i.e. lexical or post-lexical stress), the phonetic signs of glottalization may be due to prosodic prominence.

- There is phonetic evidence for a glottal stop phoneme or allophone word-finally...

    - Especially if the evidence is derived from recordings of words in isolation, the phonetic signs of glottalization can be due to phrase-final creak.

    - If the language contrasts short and long vowels, and if glottalization only appears after short vowels, glottalization may be used to ensure that the vowel remain short by closing the glottis abruptly.

Once a glottal stop phoneme/allophone is posited in a language's phonology, its realization will be subject to the same phonetic pressures found in this work, viz. prominence and

phrase-initial position. Strong prosodic positions are known to influence segmental realization, and so it is likely that segmental glottal stops will be realized more canonically as [ʔ] under prominence and in phrase-initial position. How then can one know whether a glottal stop is derived from the lexical or post-lexical phonology, if one assumes that it can lenite to ∅ in weak environments? In other words, which of the following two scenarios is right, and how do we know?

1. Deletion via lenition: /ʔ/ → ∅ when following vowel is neither prominent nor phrase-initial

2. Insertion via strengthening: ∅ → [ʔ] when initial vowel is either prominent or phrase-initial

The data from Chapter 5 show that glottal stops in English are most likely post-lexical, so neither of the above rules work; even in typically strong environments (i.e., phrase-initial position) – where segments are normally fully realized – glottal stops are not typically found unless there is also prominence. But had full glottal stops been found to always occur in *both* phrase-initial and prominent positions, it would have been hard to say whether word-initial glottal stops are phonemic (as in rule 1) or derived post-lexically (as in rule 2). In these cases, the most solid evidence for lexically-specified glottalization would come from segmental alternations (e.g., /ʔ/ stop gemination in Finnish (Suomi et al., 2008, §5.2)). In the absence of such evidence, however, researchers can appeal to phonetic clues to glottalization, bearing in mind that such clues interact closely with prosody and other factors.

## 6.2  Conclusion: glottal stop variability and invariability

To conclude, the studies of this dissertation provide insight into glottal stop variability, in terms of their production, their perception, their distribution, and (related to their distribution) their function. Despite the different sources of variability, there remains a certain invariability with respect to glottal stops.

Glottal stops can be produced with complete or incomplete closure of the glottis, and with or without ventricular incursion, which if present may also be full or incomplete. This variability appears to be random, though with further research the sources of variability may be determined. Nonetheless, full glottal stops are all characterized by the absence of vocal fold vibrations (rather than by complete glottal closure), and they tend to be more strongly produced in coda position and at the beginnings of phrases.

Glottal stops and creaky voice can be mutually confusable, especially after nasals. Listeners differ with respect to how confusable these two sounds are, presumably because, when detecting glottal stops, some listeners rely too heavily on cues shared by both creaky voice and glottal stops. Nonetheless, glottal stops and creak are less confusable in *button*-type words compared with *atlas*- and *dent*-type words, which I argue is due to the greater number of glottalization cues that may be recovered.

Although several factors are known to influence when glottal stops occur, in this study only five factors and one interaction between factors emerged as significant: prominence, phrasal position, preceding voicing irregularity, preceding pauses, the lexical category of the preceding word, and hiatus environment after a period of voicing irregularity. Nonetheless, the vast majority (95%) of full [ʔ]s in English occur in prominent and/or phrase-initial positions.

Taking into account their variability in production, perception, and distribution, the main 'glottal stop puzzles' outlined at the beginning of Chapter 1 can be answered:

- Are glottal stops really *glottal*?

  – Yes, though they usually appear with supraglottal constriction.

- Are glottal stops really *stops*?

  – Yes, they appear as stops in prosodically strong positions, especially when they are both phrase-initial and prominent.

- Why are glottal stops so common word-initially?

– Laryngealization occurs on word-initial vowels, (I argue) to enhance prominence by boosting high-frequency energy. But in especially strong positions, e.g. phrase-initially, this laryngealization is strengthened to [ʔ].

These results build on our knowledge of glottal stop production, perception, and distribution. In terms of production, glottal stops can be solely glottal, rather than glottoventricular (Esling et al., 2007), but they are usually the latter. Moreover, ventricular fold incursion is not only a means of arresting vocal fold vibration, because it occurs during onset [ʔ]s in Utterance-initial position, and (for both onset and coda [ʔ]) peaks midway during glottal stop closure. In terms of perception, previous work has shown that, when perceiving glottal stops, listeners can rely solely on F0 or amplitude dips (Hillenbrand & Houde, 1996; Pierrehumbert & Frisch, 1997), which are also potential cues to phrase-final creak. The results from this dissertation add to these findings by showing that listeners do confuse glottalization and phrase-final creak in certain circumstances, especially when cues to either gesture occurs during a coda nasal. However, the two gestures often remain distinct. Lastly, the results of this dissertation confirm previous work showing that prosody plays an important role in word-initial glottalization in English (Umeda, 1978; Pierrehumbert & Talkin, 1992; Dilley et al., 1996; Borroff, 2007). The results also indicate that prominence and phrasing are the *most important* factors to word-initial glottalization in English. In both Spanish and English, phrase-initial voicing is, all else equal, breathier – except for word-initial vowels that are prominent. Thus, word-initial glottalization is more closely linked to prominence, rather than general prosodic, strengthening of word-initial vowels.

# APPENDIX A

# Appendix for Chapter 3: Sample kymograms
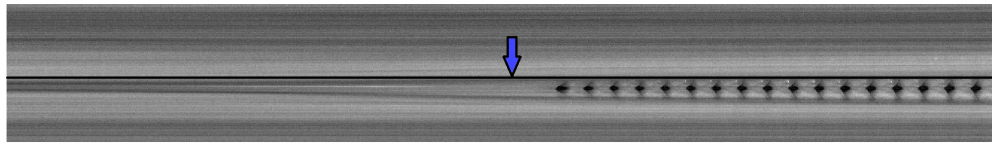


Time (frame)

(a) Onset [?]: ventricular abduction before phonation onset
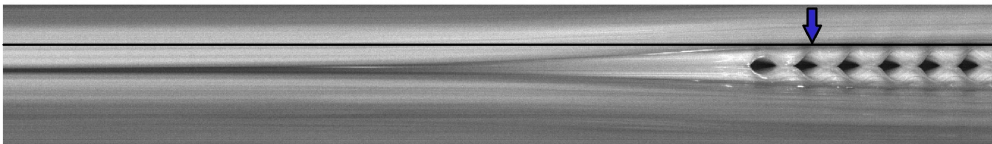


Time (frame)

(b) Onset [?]: ventricular abduction after phonation onset
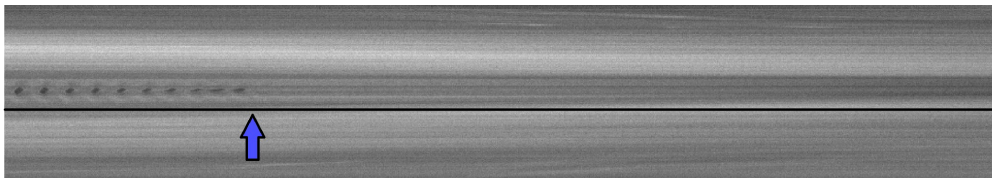


Time (frame)

(c) Coda [?]: ventricular incursion after glottal closure
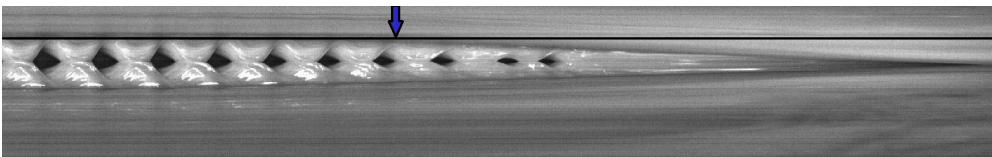


Time (frame)

(d) Coda [?]: ventricular incursion before glottal closure

Figure A.1: Examples of kymograms with different points of ventricular incursion.

# APPENDIX B

# Appendix for Chapter 4: Within-subject plots
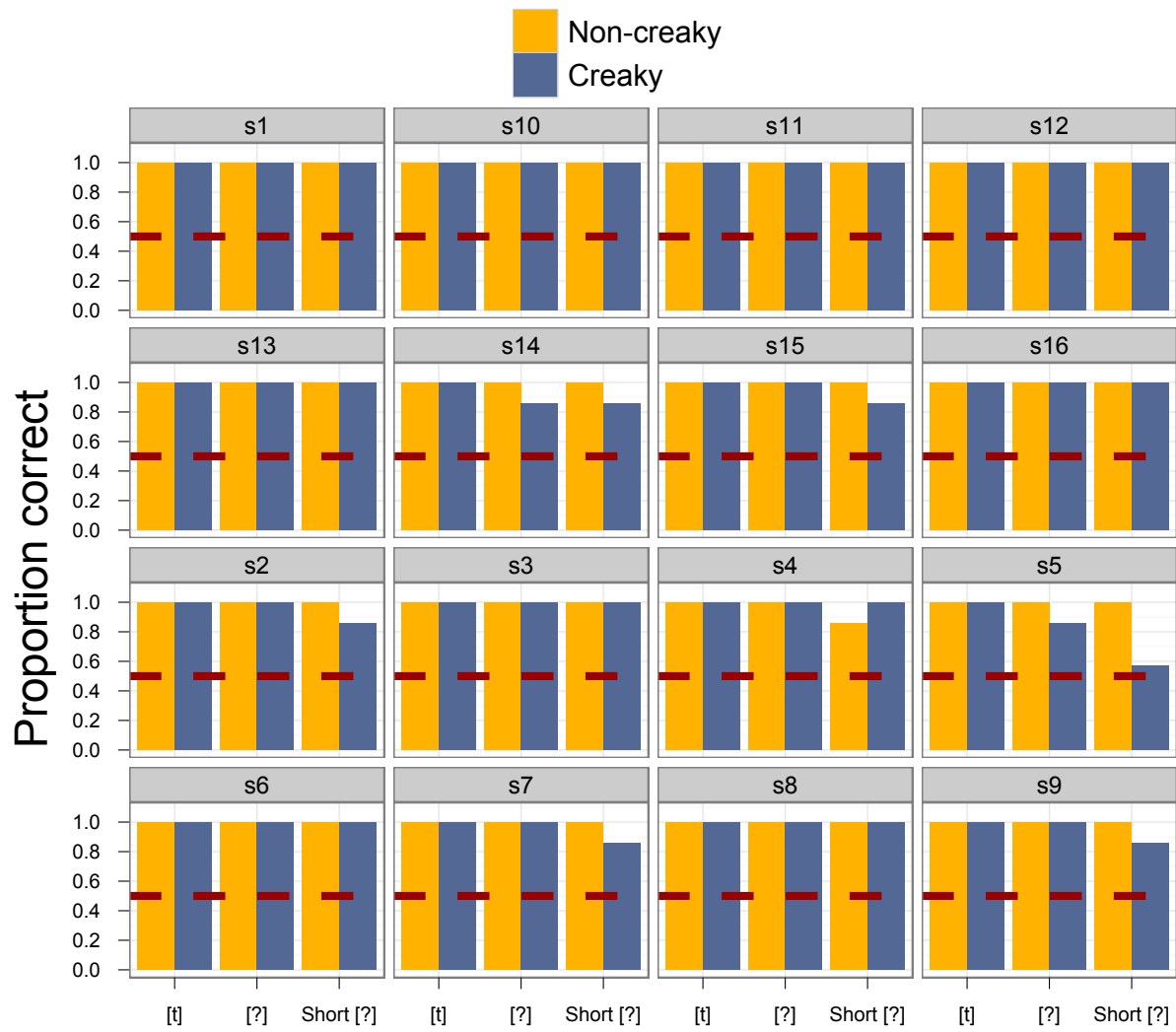


Figure B.1: Mean proportion correct for *button*-type words, separated by speaker. Chance (50%) is marked by a dashed line. '[?]' indicates [?]. For overall results, see Figure 4.2a.
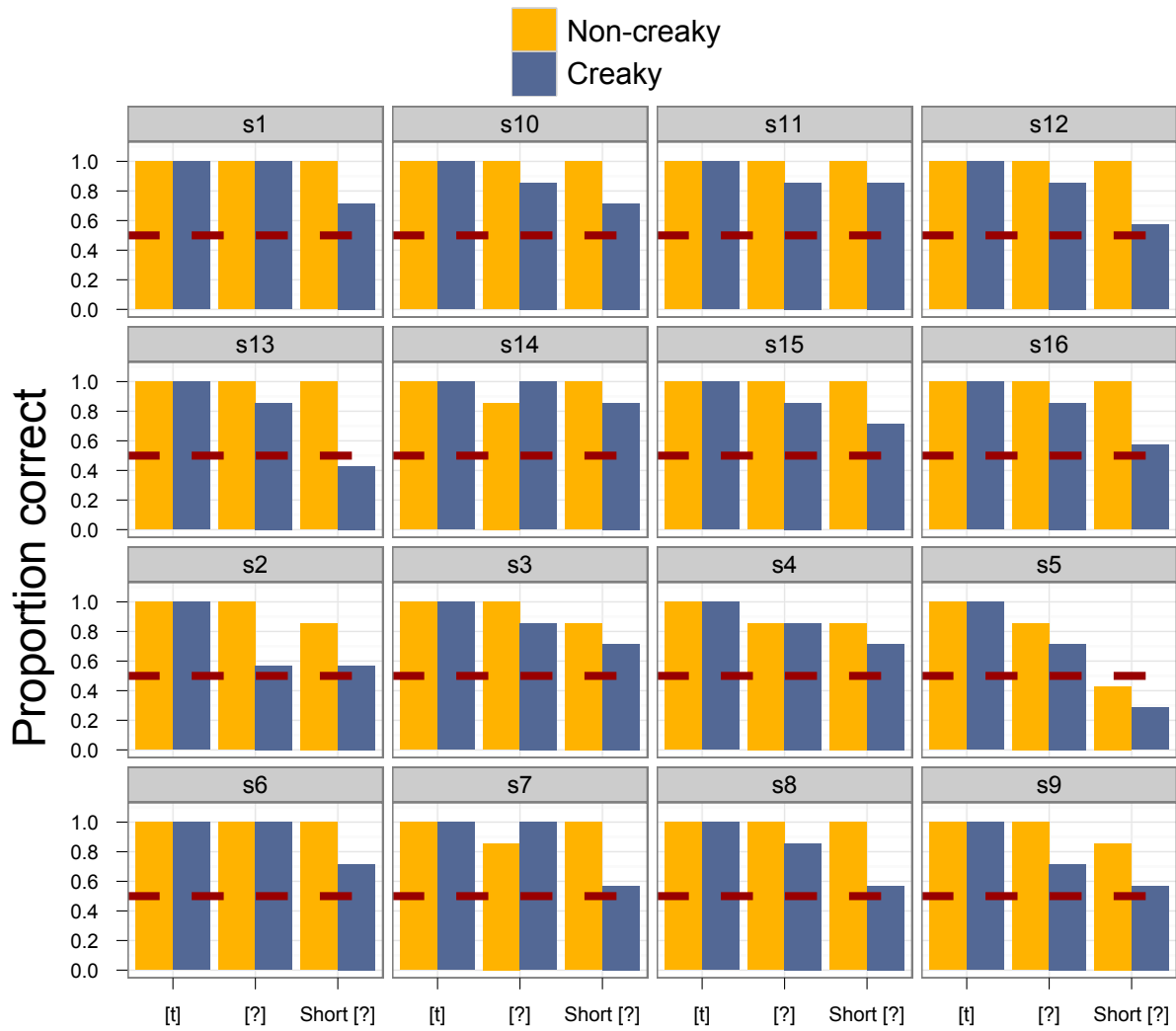
Figure B.2: Mean proportion correct for *atlas*-type words, separated by speaker. Chance (50%) is marked by a dashed line. '[ʔ]' indicates [ʔ]. For overall results, see Figure 4.3a.
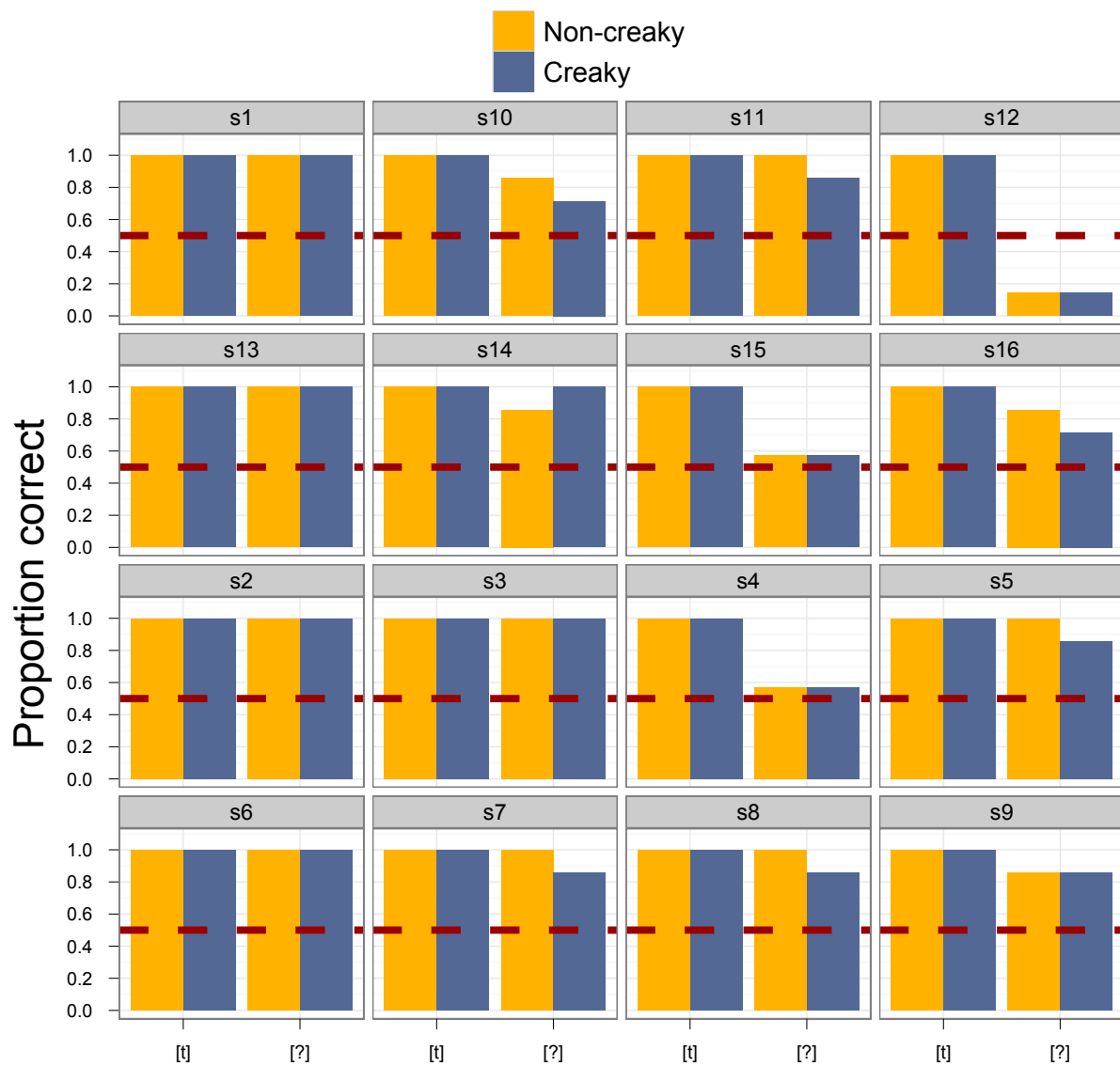
Figure B.3: Mean proportion correct for *dent*-type words, separated by speaker. Chance (50%) is marked by a dashed line. '[?]' indicates [?]. For overall results, see Figure 4.4a.
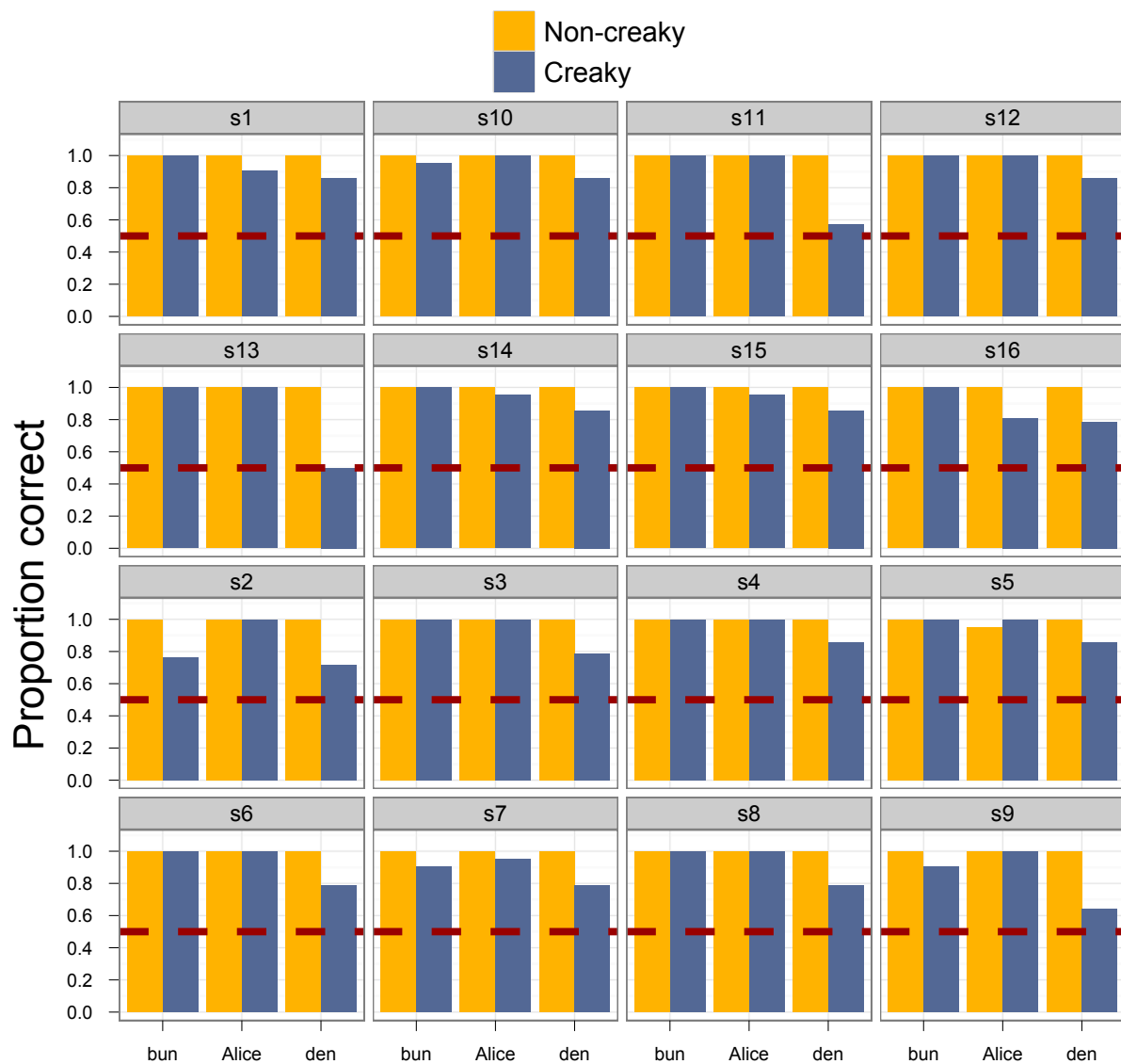
Figure B.4: Mean proportion correct for *bun*-, *Alice*-, and den-type words, separated by speaker. Chance (50%) is marked by a dashed line. For overall results, see Figure 4.5a.

# APPENDIX C

## Appendix for Chapter 4: T-tests

|          | $t$    | degrees of freedom | $p$-value |
|----------|--------|--------------------|-----------|
| Energy   | 2.42   | 13                 | < 0.05    |
| Duration | −4.12  | 17                 | < 0.001   |
| HNR05    | −3.86  | 13                 | < 0.01    |
| HNR15    | −4.18  | 13                 | < 0.01    |
| HNR25    | −5.03  | 13                 | < 0.001   |
| HNR35    | −4.31  | 13                 | < 0.001   |
| CPP      | −2.23  | 13                 | < 0.05    |

Table C.1: Significant paired t-test results for creaky segments with or without preceding glottalization.

# APPENDIX D

# Appendix for Chapter 5: Within-sonorant data
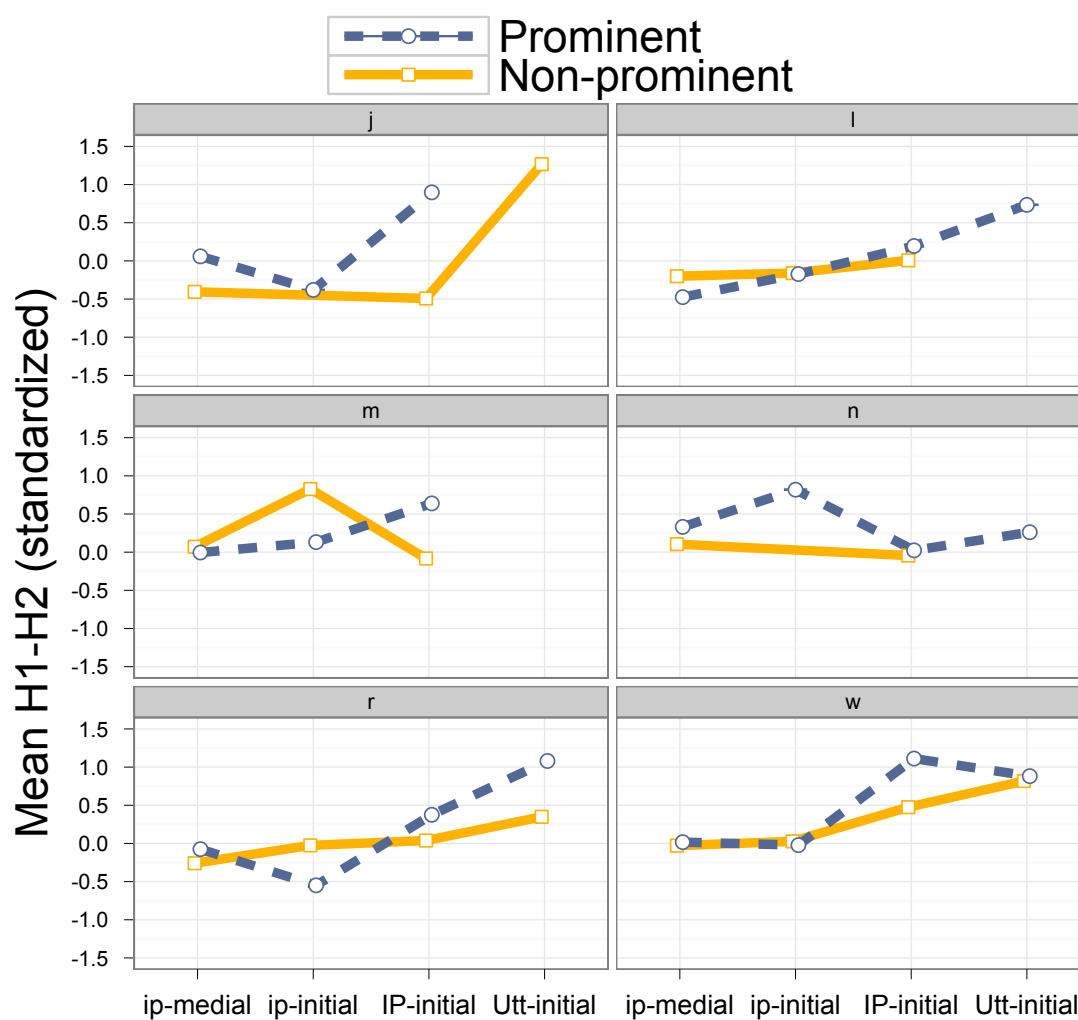


Figure D.1: Mean H1-H2 (standardized) for each sonorant analyzed in the corpus. For overall results, see Figure 5.7a.

# Bibliography

Alwan, A. (1989). Perceptual cues for place of articulation for the voiced pharyngeal and uvular consonants. *Journal of the Acoustical Society of America*, *86*, 549–556.

Andruski, J. E. (2006). Tone clarity in mixed pitch/phonation-type tones. *Journal of Phonetics*, *34*, 388–404.

Baal, B. A. B., Odden, D., & Rice, C. (2012). An analysis of North Saami gradation. *Phonology*, *29*, 165–212.

Baayen, R. H. (2008). *Analyzing Linguistic Data. A practical introduction to statistics.*. Cambridge: Cambridge University Press.

Bailly, L. (2009). *Interaction entre cordes vocales et bandes ventriculaires en phonation: Exploration in-vivo, modélisation physique, validation in-vitro (Interaction between vocal folds and ventricular folds in phonation: In-vivo exploration, physical modelling, in-vitro validation)*. Ph.D. thesis Université du Maine.

Bailly, L., Henrich, N., & Pelorson, X. (2010). Vocal fold and ventricular fold vibration in period-doubling phonation: Physiological description and aerodynamic modeling. *Journal of the Acoustical Society of America*, *127*, 3212–3222.

Baken, R. J., & Orlikoff, R. F. (2000). *Clinical Measurement of Speech and Voice*. San Diego: Singular Publishing Group.

Basbøll, H. (2003). Prosody, productivity and word structure: the stød pattern of Modern Danish. *Nordic Journal of Linguistics*, *26*, 5–44.

Bates, D., Maechler, M., & Dai, B. (2008). *lme4: Linear mixed-effects models using S4 classes*. [Computer software program] available from `http://lme4.r-forge.r-project.org/`.

Batliner, A., Burger, S., Johne, B., & Kießling, A. (1993). MÜ: A classifcation scheme for laryngealizations. In *Proceedings of ESCA workshop on prosody* (pp. 176–179). Lund.

Beckman, M. E., & Ayers Elam, G. (1997). *Guidelines for ToBI labelling, Version 3.*. The Ohio State University Research Foundation.

Belotel-Grenié, A., & Grenié, M. (1997). Types de phonation et tons en chinois standard. *Cahiers de linguistique - Asie orientale*, *26*, 249–279.

Bendor-Samuel, J. T. (1970). Some problems of segmentation in the phonological analysis of Terena. In F. R. Palmer (Ed.), *Prosodic Analysis*. London: Oxford University Press. [Reprint of article from *Word 16*, 348-355].

van den Berg, J. (1958). Myoelastic aerodynamic theory of voice production. *Journal of Speech and Hearing Research*, *1*, 227–244.

Bird, S., Caldecott, M., Campbell, F., Gick, B., & Shaw, P. A. (2008). Oral–laryngeal timing in glottalised resonants. *Journal of Phonetics*, *36*, 492–507.

Bissiri, M. P., Lecumberri, M. L., Cooke, M., & Volín, J. (2011). The role of word-initial glottal stops in recognizing English words. In *Proceedings of Interspeech 2011*. Florence, Italy.

Bissiri, M. P., & Volín, J. (2010). Prosodic structure as a predictor of glottal stops before word-initial vowels in Czech English. In *Proceedings of the 20th Czech-German Workshop on Speech Processing*.

Blankenship, B. (1997). *The time course of breathiness and laryngealization in vowels*. Ph.D. thesis UCLA.

Blankenship, B. (2002). The timing of nonmodal phonation in vowels. *Journal of Phonetics*, *30*, 163–191.

Boersma, P., & Weenink, D. (2011). Praat: doing phonetics by computer [Computer program]. Version 5.3.02. Retrieved November 10, 2011 from http://www.praat.org/.

151

Borroff, M. L. (2007). *A landmark underspecification account of the patterning of glottal stop*. Ph.D. thesis Stony Brook University.

Brunelle, M. (2009). Tone perception in Northern and Southern Vietnamese. *Journal of Phonetics*, *37*, 79–96.

Brunelle, M., Nguyễn, D. D., & Nguyễn, K. H. (2010). A laryngographic and laryngoscopic study of Northern Vietnamese tones. *Phonetica*, *67*, 147–169.

Brunner, J., & Żygis, M. (2011). Why do glottal stops and low vowels like each other? In *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 376–379).

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*, 977–990.

Buizza, E., & Plug, L. (2012). Lenition, fortition and the status of plosive affrication: the case of spontaneous RP English /t/. *Phonology*, *29*, 1–38.

Byrd, D. (1994). Relations of sex and dialect to reduction. *Speech Communication*, *15*, 39–54.

Carlson, B. F., & Esling, J. H. (2003). Phonetics and physiology of the historical shift of uvulars to pharyngeals in Nuuchahnulth (Nootka). *Journal of the International Phonetic Association*, *33*, 183–193.

Cho, T. (2005). Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /ɑ,i/ in English. *Journal of the Acoustical Society of America*, *117*, 3867–3878.

Cho, T., & Jun, S.-A. (2000). Domain-initial strengthening as enhancement of laryngeal features: Aerodynamic evidence from Korean. In J. Boyle, J.-H. Lee, & A. Okrent (Eds.), *Proceedings of Chicago Linguistic Society 36*.

Cho, T., & Keating, P. (2009). Effects of initial position versus prominence in English. *Journal of Phonetics*, *37*, 466–485.

Cohn, A. C. (1993). Nasalisation in English: phonology or phonetics. *Phonology*, *10*, 43–81.

Cole, J., Mo, Y., & Hasegawa-Johnson, M. (2010). Signal-based and expectation-based factors in the perception of prosodic prominence. *Journal of Laboratory Phonology*, *1*, 425–452.

Comrie, B. (1981). *The languages of the Soviet Union*. Carmbridge Language Surveys. Cambridge: Cambridge University Press.

Cooke, A., Ludlow, C. L., Hallett, N., & Selbie, W. S. (1997). Characteristics of vocal fold adduction related to voice onset. *Journal of Voice*, *11*, 12–22.

Davidson, L., & Erker, D. (2012). Hiatus resolution in American English: the case against glide insertion. Unpublished manuscript, NYU.

Davison, D. S. (1991). An acoustic study of so-called creaky voice in Tianjin Mandarin. *UCLA Working Papers in Phonetics*, *78*, 50–57.

de−la−Mota, C., Butragueño, P. M., & Prieto, P. (2010). Mexican Spanish intonation. In P. Prieto, & P. Roseano (Eds.), *Transcription of Intonation of the Spanish Language* (pp. 319–350). München: Lincom Europa.

Delgutte, B. (1980). Representation of speech-like sounds in the discharge patterns of auditory-nerve fibers. *Journal of the Acoustical Society of America*, *68*, 843–857.

Delgutte, B. (1982). Some correlates of phonetic distinctions at the level of the auditory nerve. In R. Carlson, & B. Granström (Eds.), *The Representation of Speech in the Peripheral Auditory System* (pp. 131–150). Amsterdam: Elsevier.

Delgutte, B., & Kiang, N. Y. S. (1984). Speech coding in the auditory nerve: I. vowel-like sounds. *Journal of the Acoustical Society of America*, *75*, 866–878.

DiCanio, C. T. (2009). The phonetics of register in Takhian Thong Chong. *Journal of the International Phonetic Association*, *39*, 162–188.

Dilley, L., Shattuck-Hufnagel, S., & Ostendorf, M. (1996). Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics*, *24*, 423–444.

Docherty, G., & Foulkes, P. (1999). Sociophonetic variation in glottals in newcastle english. In *Proceedings of the International Congress of Phonetic Sciences* (pp. 1037–1040). San Francisco.

Eddington, D., & Channer, C. (2010). American English has goʔ a loʔ of glottal stops: Social diffusion and linguistic motivation. *American Speech*, *85*, 338–351.

Eddington, D., & Savage, M. (2012). Where are the moun[ʔə]ns in Utah? *American Speech*, *87*, 336–349.

Edmondson, J. A., Chang, Y., Hsieh, F., & Huang, H. J. (2011). Reinforcing voiceless finals in Taiwanese and Hakka: Laryngoscopic case studies. In *Proceedings of the 17th International Congress of Phonetic Sciences*. Hong Kong.

Edmondson, J. A., & Esling, J. H. (2006). The valves of the throat and their functioning in tone, vocal register and stress: laryngoscopic case studies. *Phonology*, *23*, 157–191.

Edmondson, J. A., Esling, J. H., Harris, J. G., & Wei, J. (2004). A phonetic study of Sui consonants and vowels. *Mon-Khmer Studies*, *34*, 47–66.

Epstein, M. A. (2002). *Voice quality and prosody in English*. Ph.D. thesis UCLA.

Esling, J. H. (1996). Pharyngeal consonants and the aryepiglottic sphincter. *Journal of the International Phonetic Association*, *26*, 65–88.

Esling, J. H. (1999). The IPA categories "pharyngeal" and "epiglottal": Laryngoscopic observations of pharyngeal articulations and larynx height. *Language and Speech*, *42*, 349–372.

Esling, J. H. (2003). Glottal and epiglottal stop in Wakashan, Salish, and Semitic. In *Proceedings of the International Congress of Phonetic Sciences* (pp. 1707–1710). Barcelona.

Esling, J. H., Fraser, K. E., & Harris, J. G. (2005). Glottal stop, glottalized resonants, and pharyngeals: A reinterpretation with evidence from a laryngoscopic study of Nuuchahnulth (Nootka). *Journal of Phonetics*, *33*, 383–410.

Esling, J. H., & Harris, J. G. (2003). An expanded taxonomy of states of the glottis. In *Proceedings of the International Congress of Phonetic Science* (pp. 1049–1052). Barcelona.

Esling, J. H., & Harris, J. G. (2005). States of the glottis: An articulatory phonetic model based on laryngoscopic observations. In W. J. Hardcastle, & J. M. Beck (Eds.), *A Figure of Speech: a Festschrift for John Laver* (pp. 347–383). Mahwah, NJ: Erlbaum.

Esling, J. H., Zeroual, C., & Crevier-Buchman, L. (2007). A study of muscular synergies at the glottal, ventricular and aryepiglottic levels. In *Proceedings of the International Congress of Phonetic Sciences* (pp. 585–588). Saarbrücken.

Esposito, C. M. (2010). Variation in contrastive phonation in Santa Ana Del Valle Zapotec. *Journal of the International Phonetic Association*, *40*, 181–198.

Esposito, C. M. (2012). An acoustic and electroglottographic study of White Hmong phonation. *Journal of Phonetics*, *40*, 466–476.

Esposito, C. M., & Scarborough, R. (2004). An acoustic and electroglottographic study of VʔV in two indigenous American languages. *Journal of the Acoustical Society of America*, *116*, 2546.

Faaborg-Andersen, K. K. (1957). Electromyographic investigation on intrinsic laryngeal muscles in humans: an investigation of subjects with normally movable vocal cords and patients with vocal cord paresis. *Acta physiologica Scandinavica*, *41*, Suppl. 140.

Fant, G., Kruckenberg, A., & Liljencrants, J. (2000). Acoustic-phonetic analysis of prominence in Swedish. In A. Botinis (Ed.), *Intonation: analysis, modelling and technology* (pp. 55–86). Dordrecht: Kluwer Academic Publishers.

Fischer-Jørgensen, E. (1987). A phonetic study of the stød in Standard Danish. *Annual report of the Institute of Phonetics University of Copenhagen*, *21*, 55–265.

Fischer-Jørgensen, E. (1989). Phonetic analysis of the stød in Standard Danish. *Phonetica*, *46*, 1–59.

Fougeron, C. (2001). Articulatory properties of initial segments in several prosodic constituents in French. *Journal of Phonetics*, *29*, 109–135.

Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America*, *101*, 3728–3740.

Frazier, M. (2009). *The production and perception of pitch and glottalization in Yucatec Maya*. Ph.D. thesis University of North Carolina, Chapel Hill.

Freeman, E., Woo, P., Saxman, J. H., & Murry, T. (2012). A comparison of sung and spoken phonation onset gestures using high-speed digital imaging. *Journal of Voice*, *26*, 226–238.

Fujimura, O. (1990). Methods and goals of speech production research. *Language and Speech*, *33*, 195–258.

Fujimura, O., & Sawashima, M. (1971). Consonant sequences and laryngeal control. *Annual Bulletin of Research Institute of Logopedics and Phoniatrics, University of Tokyo*, *5*, 1–13.

Garellek, M. (2010). *The acoustics of coarticulated non-modal phonation*. Master's thesis UCLA.

Garellek, M. (2011). Lexical effects on English vowel laryngealization. In *Proceedings of the International Congress of Phonetic Sciences* (pp. 727–730). Hong Kong.

Garellek, M. (2012). Glottal stops before word-initial vowels in American English: distribution and acoustic characteristics. *UCLA Working Papers in Phonetics*, *110*, 1–23.

Garellek, M., & Keating, P. (2011). The acoustic consequences of phonation and tone interactions in Jalapa Mazatec. *Journal of the International Phonetic Association*, *41*, 185–205.

Garellek, M., Keating, P., Esposito, C. M., & Kreiman, J. (2013). Voice quality and tone identification in White Hmong. *Journal of the Acoustical Society of America*, *133*, 1078–1089.

Garner, W. R., & Felfoldy, G. L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology*, *1*, 225–241.

Gay, T., Hirose, H., Strome, M., & Sawashima, M. (1972). Electromyography of the intrinsic laryngeal muscles during phonation. *Annals of Otology, Rhinology, and Laryngology*, *81*, 401–409.

Gerfen, C., & Baker, K. (2005). The production and perception of laryngealized vowels in Coatzospan Mixtec. *Journal of Phonetics*, *33*, 311–334.

Gerratt, B. R., & Kreiman, J. (2001). Toward a taxonomy of nonmodal phonation. *Journal of Phonetics*, *29*, 365–381.

Gimba, A. M. (2000). *Bole verb morphology*. Ph.D. thesis UCLA.

Gordon, M., & Ladefoged, P. (2001). Phonation types: a cross-linguistic overview. *Journal of Phonetics*, *29*, 383–406.

Gray, H. (1918). *Anatomy of the Human Body*. Philadelphia: Lea and Febiger. [Published online by Bartleby.com, May 2000]. `http://www.bartleby.com/107/`. Accessed on June 4, 2011.

Guion, S. G., Post, M. W., & Payne, D. L. (2004). Phonetic correlates of tongue root vowel contrasts in maa. *Journal of Phonetics*, *32*, 517–542.

Hagen, A. (1997). *Linguistic functions of glottalizations and their language-specific use in English and German*. Master's thesis Friedrich-Alexander-Universität Erlangen-Nürnberg.

Halle, M., & Stevens, K. N. (1971). A note on laryngeal features. *Quarterly Progress Report, MIT Research Laboratory of Electronics*, *101*, 198–212.

Hanson, H. M., Stevens, K. N., Kuo, H.-K. J., Chen, M. Y., & Slifka, J. (2001). Towards models of phonation. *Journal of Phonetics*, *29*, 451–480.

Harris, J. G. (1999). States of the glottis for voiceless plosives. In *Proceedings of the International Congress of Phonetic Sciences* (pp. 2041–2044). San Francisco.

Harris, J. G. (2001). States of the glottis of Thai voiceless stops and affricates. In K. Tingsabadh, & A. S. Abramson (Eds.), *Essays in Tai linguistics* (pp. 3–11). Bangkok: Chulalongkorn University Press.

Hassan, Z. M., & Esling, J. H. (2007). Laryngoscopic (articulatory) and acoustic evidence of a prevailing emphatic feature over the word in Arabic. In *Proceedings of the International Congress of Phonetic Sciences* (pp. 1753–1756). Saarbrücken.

Hayes, B. (2009). *Introductory Phonology*. Oxford: Blackwell.

Henrich, N., d'Alessandro, C., Doval, B., & Castellengo, M. (2004). On the use of the derivative of electroglottographic signals for characterization of nonpathological phonation. *Journal of the Acoustical Society of America*, *115*, 1321–1332.

Henton, C., & Bladon, A. (1988). Creak as a sociophonetic marker. In L. M. Hyman, & C. N. Li (Eds.), *Language, Speech, and Mind: Studies in Honour of Victoria A. Fromkin* (pp. 3–29). London: Routledge.

Henton, C., Ladefoged, P., & Maddieson, I. (1992). Stops in the world's languages. *Phonetica*, *49*, 65–101.

Herbst, C., & Ternström, S. (2006). A comparison of different methods to measure the EGG contact quotient. *Logopedics Phoniatrics Vocology*, *31*, 126–138.

Higginbottom, E. (1964). Glottal reinforcement in English. *Transactions of the Philological Society*, *63*, 129–142.

Hillenbrand, J. M., & Houde, R. A. (1996). Role of F0 and amplitude in the perception of glottal stops. *Journal of Speech and Hearing Research*, *39*, 1182–1190.

Hirano, M., & Ohala, J. (1967). Use of hooked-wire electrodes for eletromyography of the intrinsic laryngeal muscles. *UCLA Working Papers in Phonetics*, *7*, 35–45.

Hirano, M., Ohala, J., & Vennard, W. (1969). The function of laryngeal muscles in regulating fundamental frequency and intensity of phonation. *Journal of Speech and Hearing Research*, *12*, 616–628.

Hirose, H. (1997). Investigating the physiology of laryngeal structures. In W. J. Hardcastle, & J. Laver (Eds.), *The Handbook of Phonetic Sciences* (pp. 116–136). Oxford: Blackwell.

Hirose, H., & Gay, T. (1973). Laryngeal control in vocal attack. *Folia Phoniatrica*, *25*, 203–213.

Holmberg, E. B., Hillman, R. E., Perkell, J. S., Guiod, P., & Goldman, S. L. (1995). Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice. *Journal of Speech and Hearing Research*, *38*, 1212–1223.

Houser, M. J., Kataoka, R., & Toosarvandani, M. (2006). Pluractional reduplication in Northern Paiute. Presented at the Friends of Uto-Aztecan Languages Conference.

Howard, D. M. (1995). Variation of electrolaryngographically derived closed quotient for trained and untrained adult female singers. *Journal of Voice*, *9*, 163–172.

Hsu, C.-S., & Jun, S.-A. (1998). Prosodic strengthening in Taiwanese: syntagmatic or paradigmatic? *UCLA Working Papers in Phonetics*, *96*, 69–89.

Huber, D. (1988). *Aspects of the communicative function of voice in text intonation: constancy and variability in Swedish fundamental frequency contours*. Ph.D. thesis University of Gothenburg.

Huffman, M. K. (2005). Segmental and prosodic effects on coda glottalization. *Journal of Phonetics*, *33*, 335–362.

Hume, E., Gett, S., Hovey, L., Scudieri, K., & Spagnol, M. (2010). Cues to word-final geminate consonants in Maltese. Poster presented at LabPhon12, Albuquerque, NM.

Hyman, L. M. (1988). The phonology of final glottal stops. In *Proceedings of the 1st Western Conference on Linguistics (WECOL1)* (pp. 113–130). Fresno: California State University.

Janhunen, J. (1986). *Glottal stop in Nenets*. Helsinki: Suomalais-Ugrilainen Seura.

Jongenburger, W., & van Heuven, V. J. J. P. (1991). The distribution of (word initial) glottal stop in Dutch. In F. Drijkoningen, & A. van Kemenade (Eds.), *Linguistics in the Netherlands* (pp. 101–110). Amsterdam: John Benjamins.

Jun, S.-A. (2005). Prosodic typology. In S.-A. Jun (Ed.), *Prosodic typology* (pp. 430–458). Oxford: Oxford University Press.

Jun, S.-A., Beckman, M., & Lee, H.-J. (1998). Fiberscopic evidence for the influence on vowel devoicing of the glottal configurations for Korean obstruents. *UCLA Working Papers in Phonetics*, *96*, 43–68.

Keating, P. (2006). Phonetic encoding of prosodic structure. In J. Harrington, & M. Tabain (Eds.), *Speech production: Models, phonetic processes, and techniques* Macquarie Monographs in Cognitive Science (pp. 167–186). Psychology Press.

Keating, P., Cho, T., Fougeron, C., & Hsu, C.-S. (2003). Domain-initial articulatory strengthening in four languages. In J. Local, R. Ogden, & R. Temple (Eds.), *Phonetic Interpretation (Papers in Laboratory Phonology 6)* (pp. 143–161). Cambridge: Cambridge University Press.

Keating, P., Esposito, C., Garellek, M., Khan, S., & Kuang, J. (2011). Phonation contrasts across languages. In *Proceedings of the International Conference of Phonetic Sciences* (pp. 1046–1049). Hong Kong.

Khan, S. (2008). *Intonational phonology and focus prosody of Bengali*. Ph.D. thesis UCLA.

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America*, *87*, 820–857.

Kochanski, G., Grabe, E., Coleman, J., & Rosner, B. (2005). Loudness predicts prominence: fundamental frequency lends little. *Journal of the Acoustical Society of America*, *118*, 1038–1054.

Kohler, K. J. (1994). Glottal stops and glottalization in German. Data and theory of connected speech processes. *Phonetica*, *51*, 38–51.

Kreiman, J. (1982). Perception of sentence and paragraph boundaries in natural conversation. *Journal of Phonetics*, *10*, 163–175.

Kreiman, J., Garellek, M., & Esposito, C. (2011). Perceptual importance of the voice source spectrum from H2 to 2 kHz. *Journal of the Acoustical Society of America*, *130*, 2570.

Kreiman, J., Shue, Y.-L., Chen, G., Iseli, M., Gerratt, B. R., Neubauer, J., & Alwan, A. (2012). Variability in the relationships among voice quality, harmonic amplitudes, open quotient, and glottal area waveform shape in sustained phonation. *Journal of the Acoustical Society of America*, *132*, 2625–2632.

Kreiman, J., & Sidtis, D. (2011). *Foundations of Voice Studies*. Oxford: Wiley-Blackwell.

Kuang, J. (2011). *Production and perception of the phonation contrast in Yi*. Master's thesis UCLA.

Kuang, J. (2012). Registers in tonal contrasts. *UCLA Working Papers in Phonetics*, *110*, 46–64.

Kuang, J., & Keating, P. (2012). Glottal articulations of phonation contrasts and their acoustic and perceptual consequences. *UCLA Working Papers in Phonetics*, *111*, 123–161.

Ladd, D. R. (1984). Declination: a review and some hypotheses. *Phonology Yearbook*, *1*, 53–74.

Ladd, D. R. (1988). Declination "reset" and the hierarchical organization of utterances. *Journal of the Acoustical Society of America*, *84*, 530–544.

Ladd, D. R. (2008). *Intonational Phonology*. (2nd ed.). Cambridge: Cambridge University Press.

Ladefoged, P. (1971). *Preliminaries to linguistic phonetics*. Chicago: University of Chicago.

Ladefoged, P., & Maddieson, I. (1996). *The Sounds of the World's Languages*. Oxford: Blackwell.

Laufer, A., & Condax, I. D. (1981). The function of the epiglottis in speech. *Language and Speech*, *24*, 39–62.

Laver, J. (1980). *The Phonetic Description of Voice Quality*. Cambridge: Cambridge University Press.

Laver, J. (1994). *Principles of phonetics*. Cambridge: Cambridge University Press.

Lehiste, I. (1972). Some observations concerning the third tone in Latvian. In A. Valdman (Ed.), *Papers in linguistics and phonetics to the memory of Pierre Delattre* (pp. 309–315). The Hague: Mouton.

Lichtman, K., Chang, S., Cramer, J., del Rio, C. C., Hallett, J., Huensch, A., & Morales, A. (2010). IPA illustration of Q'anjob'al. *Studies in Linguistic Sciences: Illinois Working Papers*, (pp. 1–23).

Lindqvist, J. (1972). Laryngeal articulation studied on Swedish subjects. *STL-QPSR*, *2-3*, 10–27.

Local, J., Wells, W. H. G., & Sebba, M. (1985). Phonology for conversation: Phonetic aspects of turn delimitation in London Jamaican. *Journal of Pragmatics*, *9*, 309–330.

Local, J. K., Kelly, J., & Wells, W. H. G. (1986). Towards a phonology of conversation: turn-taking in Tyneside English. *Journal of Linguistics*, *22*, 411–437.

Lombardi, L. (1991). *Laryngeal features and laryngeal neutralization*. Ph.D. thesis University of Massachusetts.

Lombardi, L. (2002). Coronal epenthesis and markedness. *Phonology*, *19*, 219–251.

Ludlow, C. L., Sedory, S. E., & Fujita, M. (1991). Neurophysiological control of vocal fold adduction and abduction for phonation onset and offset during speech. In J. Gauffin, & B. Hammarberg (Eds.), *Vocal Fold Physiology. Acoustics, Perceptual, and Physiological Aspects of Voice Mechanisms* (pp. 197–205). San Diego: Singular Publishing Group.

Maddieson, I. (1984). *Patterns of Sounds*. Cambridge: Cambridge University Press.

Maddieson, I., Flavier, S., Marsico, E., & Pellegrino, F. (2011). LAPSyD: Lyon-Albuquerque Phonological Systems Databases, Version 1.0. `http://www.lapsyd.ddl.ish-lyon.cnrs.fr/`. Last checked on April 26, 2013.

Maddieson, I., & Precoda, K. (1990). Updating UPSID. *UCLA Working Papers in Phonetics*, *74*, 104–111. [UPSID searched online via `http://web.phonetik.uni-frankfurt.de/Upsid.html`, June 4 2011].

Malécot, A. (1975). The glottal stop in French. *Phonetica*, *31*, 51–63.

Mathisen, A. G. (1999). Sandwell, West Midlands: Ambiguous perspectives on gender patterns and models of change. In P. Foulkes, & G. Docherty (Eds.), *Urban Voices: Accent Studies in the British Isles* (pp. 107–123). London: Arnold.

McCarthy, J. J. (1994). The phonetics and phonology of Semitic pharyngeals. In P. A. Keating (Ed.), *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III* (pp. 191–233). Cambridge: Cambridge University Press.

Mees, I., & Collins, B. (1999). Cardiff: a real-time study of glottalisation. In P. Foulkes, & G. Docherty (Eds.), *Urban Voices: Accent Studies in the British Isles* (pp. 85–202). London: Arnold.

Mendelsohn, A. H., & Zhang, Z. (2011). Phonation threshold pressure and onset frequency in a two-layer physical model of the vocal folds. *Journal of the Acoustical Society of America*, *130*, 2961–2968.

Michaud, A. (2004). Final consonants and glottalization: New perspectives from Hanoi Vietnamese. *Phonetica*, *61*, 119–146.

Miller, A. L. (2007). Guttural vowels and guttural co-articulation in Ju|'hoansi. *Journal of Phonetics*, *35*, 56–84.

Milroy, J., Milroy, L., Hartley, S., & Walshaw, D. (1994). Glottal stops and Tyneside glottalization: Competing patterns of variation and change in British English. *Language Variation and Change*, *6*, 327–357.

Moisik, S., & Esling, J. (2012). Modeling the mechanical influence of epilaryngeal tube stricture on vocal fold dynamics. Presented at the 2012 International Conference on Vocal Fold Physiology and Biomechanics, Erlangen.

Moisik, S. R., & Esling, J. H. (2011). The 'whole larynx' approach to laryngeal features. In *Proceedings of the 17th International Congress of Phonetic Sciences*. Hong Kong.

Mompeán, J., & Gómez, F. A. (2011). Hiatus resolution strategies in non-rhotic English: the case of /r/-liaison. In *Proceedings of the International Conference of Phonetic Sciences* (pp. 1414–1417). Hong Kong.

Moore, P. (1938). Motion picture studies of the vocal folds and vocal attack. *Journal of Speech Disorders*, *3*, 235–238.

Myers, S. (2012). Final devoicing: Production and perception studies. In T. Borowsky, S. Kawahara, & M. Sugahara (Eds.), *Prosody Matters: Essays in honor of Elisabeth Selkirk* (pp. 148–180). London: Equinox Publishing.

Nakatani, L. H., & Dukes, K. D. (1977). Locus of segmental cues for word juncture. *Journal of the Acoustical Society of America*, *62*, 714–719.

Nguyen, V. L., & Edmondson, J. A. (1998). Tones and voice quality in modern northern Vietnamese: Instrumental case studies. *Mon-Khmer Studies*, *28*, 1–18.

Ogden, R. (2001). Turn transition, creak and glottal stop in Finnish talk-in-interaction. *Journal of the International Phonetic Association*, *31*, 139–152.

Orlikoff, R. F. (1991). Assessment of the dynamics of vocal fold contact from the electroglottogram. *Journal of Speech and Hearing Research*, *34*, 1066–1072.

Orlikoff, R. F., Deliyski, D. D., Baken, R. J., & Watson, B. C. (2009). Validation of a glottographic measure of glottal attack. *Journal of Voice*, *23*, 164–168.

O'Shaughnessy, D., & Allen, J. (1983). Linguistic modality effects on fundamental frequency in speech. *Journal of the Acoustical Society of America*, *74*, 1155–1171.

Ostendorf, M., Price, P. J., & Shattuck-Hufnagel, S. (1995). *The Boston University radio news corpus*. Technical Report ECS-95-001 Boston University.

Pham, A. H. (2003). *Vietnamese tone: a new analysis*. New York: Routledge.

Pierrehumbert, J. (1995). Prosodic effects on glottal allophones. In O. Fujimura, & M. Hirano (Eds.), *Vocal fold physiology: voice quality control* (pp. 39–60). San Diego: Singular Publishing Group.

Pierrehumbert, J., & Talkin, D. (1992). Lenition of /h/ and glottal stop. In G. J. Docherty, & D. R. Ladd (Eds.), *Papers in Laboratory Phonology II* (pp. 90–117). Cambridge: Cambridge University Press.

Pierrehumbert, J. B., & Frisch, S. (1997). Synthesizing allophonic glottalization. In J. P. Van Santen, R. W. Sproat, J. P. Olive, & J. Hirschberg (Eds.), *Progress in speech synthesis* (pp. 9–26). New York: Springer-Verlag volume 1.

Pierrehumbert, J. B., & Hirschberg, J. (1990). The meaning of intonation contours in the interpretation of discourse. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Plans and Intentions in Communication and Discourse* SDF Benchmark Series in Computational Linguistics (pp. 271–311). Cambridge: MIT Press.

Podesva, R. J., & Lee, S. (2010). Voice quality variation and gender in Washington, DC. Presented at New Ways of Analyzing Variation, San Antonio, TX.

Pompino-Marschall, B., & Żygis, M. (2011). Glottal marking of vowel-initial words in German. In *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 1626–1629). Hong Kong.

Price, P. J., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America*, *90*, 2956–2970.

Priestly, T. M. (1976). A note on the glottal stop. *Phonetica*, *33*, 268–274.

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria.

Ratliff, M. (1992). *Meaningful Tone: A study of tonal morphology in compounds, form classes and expressive phrases in White Hmong*. Monograph Series on Southeast Asia. DeKalb, IL: Northern Illinois University, Center for Southeast Asian Studies.

Redi, L., & Shattuck-Hufnagel, S. (2001). Variation in the realization of glottalization in normal speakers. *Journal of Phonetics*, *29*, 407–429.

Reidenbach, M. M. (1998). The muscular tissue of the vestibular folds of the larynx. *European Archives of Oto-Rhino-Laryngology*, *255*, 365–367.

Roach, P. J. (1973). Glottalization of English /p/, /t/, /k/ and /tʃ/ – a re-examination. *Journal of the International Phonetic Association*, *3*, 10–21.

Roach, P. J. (1979). Laryngeal-oral coarticulation in glottalized English plosives. *Journal of the International Phonetic Association*, *9*, 2–6.

van Rossum, M. A., van As-Brooks, C. J., Hilgers, F. J. M., & Roozen, M. (2009). Quality of 'glottal' stops in tracheoesophageal speakers. *Clinical Linguistics and Phonetics*, *23*, 1–14.

Rothenberg, M., & Mashie, J. J. (1988). Monitoring vocal fold abduction through vocal fold contact area. *Journal of Speech and Hearing Research*, *31*, 338–351.

Sakakibara, K.-I., Kimura, M., Imagawa, H., Niimi, S., & Tayama, N. (2004). Physiological study of the supraglottal structure. In *Proceedings of the International Conference on Voice Physiology and Biomechanics*. Marseilles.

Seid, H., Yegnanarayana, B., & Rajendran, S. (2012). Spotting glottal stop in Amharic in continuous speech. *Computer Speech and Language*, *26*, 293–305.

Shue, Y.-L., Chen, G., & Alwan, A. (2010a). On the interdependencies between voice quality, glottal gaps, and voice-source related acoustic measures. In *Proceedings of Interspeech 2010* (pp. 34–37).

Shue, Y.-L., Keating, P. A., Vicenik, C., & Yu, K. (2011). VoiceSauce: A program for voice analysis. In *Proceedings of the International Congress of Phonetic Sciences* (pp. 1846–1849). Hong Kong.

Shue, Y.-L., Shattuck-Hufnagel, S., Iseli, M., Jun, S.-A., Veilleux, N., & Alwan, A. (2010b). On the acoustic correlates of high and low nuclear pitch accents in American English. *Speech Communication*, *52*, 106–122.

Silverman, D. (1995). *Phasing and recoverability*. Ph.D. thesis UCLA. [Published in 1997 by Garland].

Silverman, D. (2003). Pitch discrimination between breathy vs. modal phonation. In J. Local, R. Ogden, & R. Temple (Eds.), *Phonetic Interpretation (Papers in Laboratory Phonology 6)* (pp. 293–304). Cambridge: Cambridge University Press.

Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., & Hirschberg, J. (1992). ToBI: A standard for labeling English prosody. In *Proceedings of the Second International Conference on Spoken Language Processing* (pp. 867–870). Banff, Alberta.

Slifka, J. (2000). *Respiratory constraints on speech production at prosodic boundaries*. Ph.D. thesis MIT.

Slifka, J. (2006). Some physiological correlates to regular and irregular phonation at the end of an utterance. *Journal of Voice*, *20*, 171–186.

Stevens, K. N. (1977). Physics of laryngeal behavior and larynx modes. *Phonetica*, *34*, 264–279.

Stevens, K. N. (2000). *Acoustic Phonetics*. Cambridge: MIT Press.

Straka, G. (1963). La division des sons du langage en voyelles et consonnes peut-elle être justifiée? *Travaux de linguistique et de littérature, Université de Strasbourg*, *1*, 17–99.

Sumner, M., & Samuel, A. G. (2005). Perception and representation of regular variation: The case of final /t/. *Journal of Memory and Language*, *52*, 322–338.

Suomi, K., Toivanen, J., & Ylitalo, R. (2008). *Finnish Sound Structure. Phonetics, phonology, phonotactics and prosody*. Oulo, Finland: Oulu University Press.

Surana, K. K. (2006). *Classification of vocal fold vibration as regular or irregular in normal, voiced speech*. Master's thesis MIT.

Švec, J. G., & Schutte, H. K. (1996). Videokymography: High-speed line scanning of vocal fold vibration. *Journal of Voice*, *10*, 201–205.

Thornes, T. J. (2003). *A Northern Paiute grammar: with texts*. Ph.D. thesis University of Oregon.

Titze, I. R. (1992). Phonation threshold pressure: A missing link in glottal aerodynamics. *Journal of the Acoustical Society of America*, *91*, 2926–2935.

Turk, A. E., & Sawusch, J. R. (1996). The processing of duration and intensity cues to prominence. *Journal of the Acoustical Society of America*, *99*, 3782–3790.

Umeda, N. (1978). Occurrence of glottal stops in fluent speech. *Journal of the Acoustical Society of America*, *64*, 88–94.

Vance, T. J. (1977). Tonal distinctions in Cantonese. *Phonetica*, *34*, 93–107.

Watt, D., & Milroy, L. (1999). Patterns of variation and change in three Tyneside vowels: Is this dialect levelling? In P. Foulkes, & G. Docherty (Eds.), *Urban Voices: Accent Studies in the British Isles* (pp. 25–46). London: Arnold.

Wayland, R., & Jongman, A. (2003). Acoustic correlates of breathy and clear vowels: The case of Khmer. *Journal of Phonetics*, *31*, 181–201.

Wells, J. C. (1982). *Accents of English 2: The British Isles*. Cambridge: Cambridge University Press.

Welmers, W. E. (1973). *African Language Structures*. Berkeley and Los Angeles: University of California Press.

Westbury, J. R., & Keating, P. A. (1986). On the naturalness of stop consonant voicing. *Journal of Linguistics*, *22*, 145–166.

Wittenberg, T., Tigges, M., Mergell, P., & Eysholdt, U. (2000). Functional imaging of vocal fold vibration: digital multislice high-speed kymography. *Journal of Voice*, *14*, 422–442.

Wolk, L., Abdelli-Beruh, N. B., & Slavin, D. (2012). Habitual use of vocal fry in young adult female speakers. *Journal of Voice*, *26*, e111–e116.

Yu, K. M. (2010). Laryngealization and features for Chinese tonal recognition. In *Proceedings of Interspeech 2010*.

Yu, K. M., & Lam, H. W. (2011). The role of creaky voice in Cantonese tonal perception. In *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 2240–2243).

Yuasa, I. P. (2010). Creaky voice: a new feminine voice quality for young urban-oriented upwardly mobile American women? *American Speech*, *85*, 315–337.

Zawaydeh, B. A. (1999). *The phonetics and phonology of gutturals in Arabic*. Ph.D. thesis Indiana University.

Zemlin, W. R. (1988). *Speech and Hearing: Anatomy and Physiology*. (3rd ed.). Englewood Cliffs: Prentice Hall Inc.

Zhang, Z. (2011). Restraining mechanisms in regulating glottal closure during phonation. *Journal of the Acoustical Society of America*, *130*, 4010–4019.

Zhang, Z., Kreiman, J., & Gerratt, B. R. (2011). Perceptual sensitivity to changes in vocal fold geometry and stiffness. *Journal of the Acoustical Society of America*, *129*, 2529.