UNIVERSITY OF CALIFORNIA

Los Angeles

Voice Quality and Prosody in English

A dissertation submitted in partial satisfaction

of the requirements for the degree Doctor of Philosophy

in Linguistics

by

Melissa Ann Epstein

2002

The dissertation of Melissa Ann Epstein is approved.

Sun-Ah Jun

Peter Ladefoged

Patricia Keating, Committee Co-chair

Jody Kreiman, Committee Co-chair

University of California, Los Angeles

2002

To my parents.

Table of Contents

List of Figures

List of Tables

Acknowledgments

You may tire of reality, but you never tire of dreams.

L. M. Montgomery (1874 – 1942)

For as long as I can remember, I have always dreamed of getting a PhD. And now that it has finally happened, I would like to thank those who have both helped me keep sight of that dream and made the day-to-day realities of dream-chasing much more manageable.

First, I would like to thank my dissertation committee members: Pat Keating, Jody Kreiman, Sun-Ah Jun and Peter Ladefoged. Pat, thank you for always helping me locate the proverbial forest and thin out a few trees. Thank you also for gently guiding me all the way through grad school – you have been an amazing teacher and advisor. Jody, thanks for introducing me to the wonders and mysteries of voice quality that have occupied most of my time in grad school. Also, I would like to thank you for always reminding me that this is my *primum opus* and not my *magnum opus*. And thank you for sharing with me your love of cool socks and fine tea. Sun-Ah, thank you for long talks on intonation and life in general. And thank you for paying attention to details, particularly when I could not see them any more. Peter, thank you for helping me out with data analysis techniques and introducing me to the joys of anatomy and physiology. Thank you also for helping to produce Figure 13 of true and differentiated glottal flow. I would also like to thank Carson Schütze for working with me for the first half of this dissertation, and for rearranging his schedule to come to my orals.

I also owe a great debt to the people who first introduced me to phonetics. To Jan Van der Spiegel and his students, for teaching an avowed technophobe UNIX and basic

digital signal processing. To Mary Beckman, for giving me my first phonetics lab experience and sending me down this road, and for continuing to advise me all the way. And to Keith Johnson and Mary Beckman, for letting me know that Ohio State was always my home away from home. I am also indebted to Ben Munson, whom I met that first summer at OSU, and who insisted I would be qualified for the post-doc.

I would like to thank all the labbies over the years, for making the lab the wonderful place that it is: Adam Albright, Tim Arbisi-Kelm, Heriberto Avelino, Marco Baroni, Roger Billerey-Mosier, Heidi Fleischhacker, Bruce Hayes, Amanda Jones, Jongho Jun, Sahyang Kim, Jenny Ladefoged, Hyuck-Joon Lee, Ying Lin, Ian Maddieson, Donca Steriade, Kim Thomas-Vilakati, Moto Ueyama, Colin Wilson, Jie Zhang and Kie Ross Zuraw. In particular, I would like to thank Amy Schaeffer and Siri Tuttle for providing sage advice and experience; Mary Baltazani for long talks about intonation; Matthew Gordon for introducing me to Zapotec; Taehong Cho, Victoria Anderson and Christina Foreman for providing the late night model; Kuniko Yasu Nielsen (and Jon!) for listening; Barbara Blankenship for inadvertently getting me addicted to voice quality when she hosted me; and Narineh Hacopian for endlessly revising the dissection manual with me. I would also like to thank my classmates, who have survived with me, with a good deal of laughter and a few tears, for the past five years: Leston Buell, Ivano Caponigro, Heidi Fleischhacker, Alex MacBride, Masangu Matondo, Gianluca Storto and Harold Torrence.

I would like to thank all the Glottal Bureaucrats, with whom I have spent half my time at UCLA: Norma Antoñanzas-Barroso, David Berry, Brian Gabelman, Bruce

Gerratt, Jody Kreiman and Blas Payri. Thank you all for your endless technical assistance and advice.

I would like to thank Henry Tehrani for solving all problems in the phonetics lab at a moment's notice. I would also like to thank Christina Esposito and Rebecca Brown for labeling my corpus, bearing with me as I developed my labeling system, and waiting uncomplainingly to get paid. Thank you also to Andrea McColl for helping analyze the corpus for the original pilot study. I am also indebted to Michael Mitchell and his crew at the statistical consulting desk for advice on statistics and the inner workings of SPSS. And to Patrick Manalastas and Natasha Levy (and all the other office staff) for making the administrative bits as pain-free as possible.

There were also many, many people who have provided "non-technical" assistance over the years. To all of my friends in LA, for dragging me away from my desk and making me play. To Rachel Miller, for reading fellowship applications, water color on Sunday nights and being there even when she moved away. To Rebecca Klempner, for reading more drafts of this dissertation than anyone else and for always providing me with a hot meal, a cup of tea and a place to crash. And to Daniel Klempner, for graciously loaning me his wife. And finally, to my roommate Shoshana Levin, for endless talks about TA-ing, advisors, statistics, and, of course, sound waves. For the past three years you have played the parts of roommate, mother and sister as needed, but your best role has been that of friend.

Finally, I would like to thank my family. To my sister, Dana, for listening, laughing and coming to visit. And for giving up her bed when I've come to visit her! And to my

parents. I would like to thank my father for paving the path. Thank you for always advising me along the way, and understanding what it's like to be in grad school And to my mother, for making sure that path was always smoothly paved. Thank you for listening to me long-distance about just about everything, and for putting up with a dissertating student for the second time. Thank you both for knowing that I would survive this even when I did not – and for ensuring that I did.

תושלב״ע

VITA

August 23, 1974	Born, Rochester, New York
1995	Phi Beta Kappa University of Pennsylvania
1996	B.A., Linguistics and Asian and Middle Eastern Studies Summa Cum Laude University of Pennsylvania Philadelphia, Pennsylvania
1997	Diploma in Jewish Studies Oxford Centre for Hebrew and Jewish Studies Oxford University Oxford, United Kingdom
1997, 1998, 2000	Pre-doctoral Traineeship in Phonetics and Speech Department of Linguistics University of California, Los Angeles
1999-00	Ursula Mandel Fellowship University of California, Los Angeles
2000	M.A., Linguistics University of California, Los Angeles Los Angeles, California
2000-02	Speech Communication Technical Council Representative and Vice-chair Student Council, Acoustical Society of America
2001-02	Teaching Assistant Department of Linguistics University of California, Los Angeles

PUBLICATIONS AND PRESENTATIONS

Epstein, M.A. 1999. On the singular indefinite article in English. *UCLA Working Papers in Syntax* 2, 14-58.

_____ 2000. A comparison of linguistic and pathological breathiness using the LF model. *Journal of the Acoustical Society of America* 107, 2906.

_____ 2000. All the sounds of all the world's languages. *UCLA Working Papers in Phonetics* 99, 1-3.

_____ 2001. Measuring voice quality in dynamic speech. *Journal of the Acoustical Society of America* 109, 2413.

_____ In press. Voice quality variations in English sentences," *Journal of the Acoustical Society of America*

_____, B. Gabelman, N. Antoñanzas-Barroso, B. Gerratt and J. Kreiman. 1999. Source model adequacy for pathological voice synthesis. *Proceedings of the XIVth International Congress of Phonetics Science (ICPhS)* 99 (3), 2049-2052.

_____, N. Hacopian, and P. Ladefoged (Eds.). In press. *Dissection of the Vocal Organs: A Guide for Speech Scientists*. UCLA Working Papers in Phonetics: Los Angeles.

_____ and P. Ladefoged 2001. Phonation types in Amerindian languages. *Proceedings of the Athabaskan Languages Conference 2001*.

_____ and B. Payri 2001. The effects of vowel quality and pitch on spectral and glottal flow measurements of the voice source. *Journal of the Acoustical Society of America* 109, 2413.

Kreiman, J., B.R. Gerratt, and M. Epstein. 2000. Effects of listener experience on measures of voice quality. *Journal of the Acoustical Society of America* 108, 2532.

_____, B.R. Gerratt, B. Gabelman, N. Antoñanzas-Barroso, and M. Epstein. 2000. User's Guide: Analysis and Synthesis of Pathological Voice Quality. Unpublished technical report. UCLA.

ABSTRACT OF THE DISSERTATION

Voice Quality and Prosody in English

by

Melissa Ann Epstein

Doctor of Philosophy in Linguistics

University of California, Los Angeles, 2002

Professor Patricia Keating, Co-chair

Professor Jody Kreiman, Co-chair

This study examines the effects of prosody on voice quality variations in American English sentences. Research on voice quality variations in English requires both taking into account the various linguistic sources for changes in voice quality and adequately tracking and measuring the range of naturally occurring voice qualities in English. For instance, several consonants in English have glottal stop allophones that can cause glottalization of the preceding vowel, and there are personal and regional variations in voice quality. In this study, these effects are controlled by obtaining an idiosyncratic baseline value for each measurement for each word of the corpus. Furthermore, voice quality variations are tracked using both quantitative measurements (derived from the Liljencrants/Fant (LF) model of the glottal source) and a qualitative description of the waveforms to capture the range of voice qualities in the corpus. Modeled source pulses

were evaluated to calculate the measurements OQ (open quotient), RK (glottal skew), EE (spectral intensity) and Linearity (spectral linearity).

Samples were taken from 3 speakers producing variants of a single sentence. Tokens included interrogative and declarative sentences, and sentences with phrase-initial and phrase-final prominent words. Results show that there are consistent effects of prominence and phrase boundaries on voice quality, both across speakers and among measurements. Both prominent words and phrase-initial words displayed a "tenser" voice quality than their non-prominent and phrase-final counterparts. A tense voice quality is associated in theory with greater compression of the vocal fold and greater force of closure of the arytenoids. Acoustically, tense voice quality is correlated with low values of open quotient and glottal skew, and high values of spectral intensity and spectral linearity. On the other hand, changes in tense or lax voice quality were not correlated with changes in pitch, nor were there any effects of pitch accent type or boundary tone on modal voice quality. There was, however, an increase in non-modal phonation associated with Low boundary tones, but not with low phonetic pitch in general.

Chapter 1: Introduction

Modern speech synthesis techniques produce speech that is intelligible, but is often unpleasant to listen to for extended periods of time and over the course of many sentences. One of the missing elements in synthetic speech is an accurate representation of prosody, the changes in the speech signal that indicate the grouping and relative prominence of the components of that speech signal. Most studies on prosody have focused on changes in pitch and duration, but there are many other variables, acoustic and other, that cue listeners to the prosodic structure of the speech signal. Examples of other prosodic cues are the degree of coarticulation or lenition of consonants and vowels, visual signals such as head nodding and eyebrow raising, and changes in voice quality. The focus of this study will be the effects of prosodic structure on voice quality in American English.

In brief, voice quality is the laryngeal setting that co-occurs with the oral articulation of a voiced vowel or consonant. Voice quality can range from the stiff vocal folds of "laryngealized" or "creaky" voice, to the "normal" vocal fold vibration patterns in "modal" phonation, to the relaxed and incomplete closure of the vocal folds in "breathy" or "weak" voice (Ladefoged 1971; Ladefoged 1973; Ladefoged & Maddieson 1996).

A study of voice quality variations in any language that does not use voice quality contrastively, such as English, needs to account for and separate out the multiple linguistic factors that may cause a particular voice quality to occur. For example, the sentence "I saw a *cat*" is likely to end with a creaky voice, as can be seen in the irregular vertical striations in the spectrogram in Figure 1.

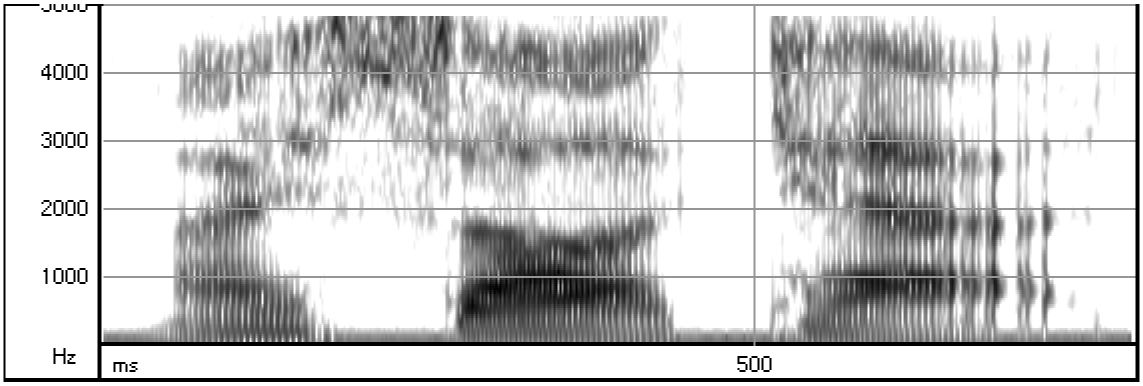


Figure 1. Spectrogram of the sentence *I saw a cat.* spoken with creaky voice on the final word.

There are a number of possibilities for the cause of this creak:

1. The utterance was a complete sentence and the ends of sentences tend be creaky.
2. The sentence was a declarative (as opposed to a interrogative) and the ends of declaratives tend to be creaky.
3. Declaratives end with a low pitch and low pitches tend to be creaky.
4. The sentence ended with the word *cat*, [kæ̰ʔt] is in allophonic variation with [kæt].
5. The word *cat* was accented. Accentuation may encourage the glottalization of the /t/ in /kæt/.

To find out why this sentence ended with a creaky voice it is necessary to have more information about the following:

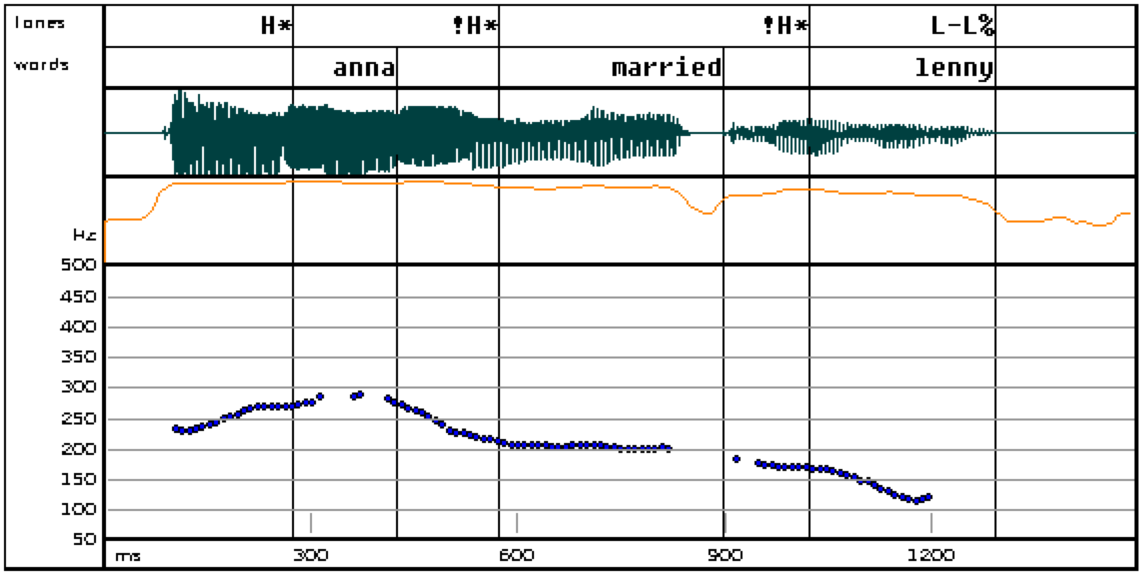
1. The effects of small linguistic variations on the voice quality of an utterance. Would the sentences "*I* saw a cat." or "I saw a *cat*?" also exhibit creaky voice at the end of the utterance, or would these sentences have creaky voice elsewhere, or not at all?
2. The voice quality tendencies of the person speaking the sentence. Does this person tend to pronounce *cat* as [kæ̰ʔt]? Or, would she show the same effects if she had an overall creaky voice?

In summary, there are three goals for this study. The first is to identify the prosodic locations for voice quality changes. The second is to identify the voice quality used by speakers in these locations. The third goal is to develop a measurement system that factors out the individual voice qualities of the speakers and the segmental voice qualities

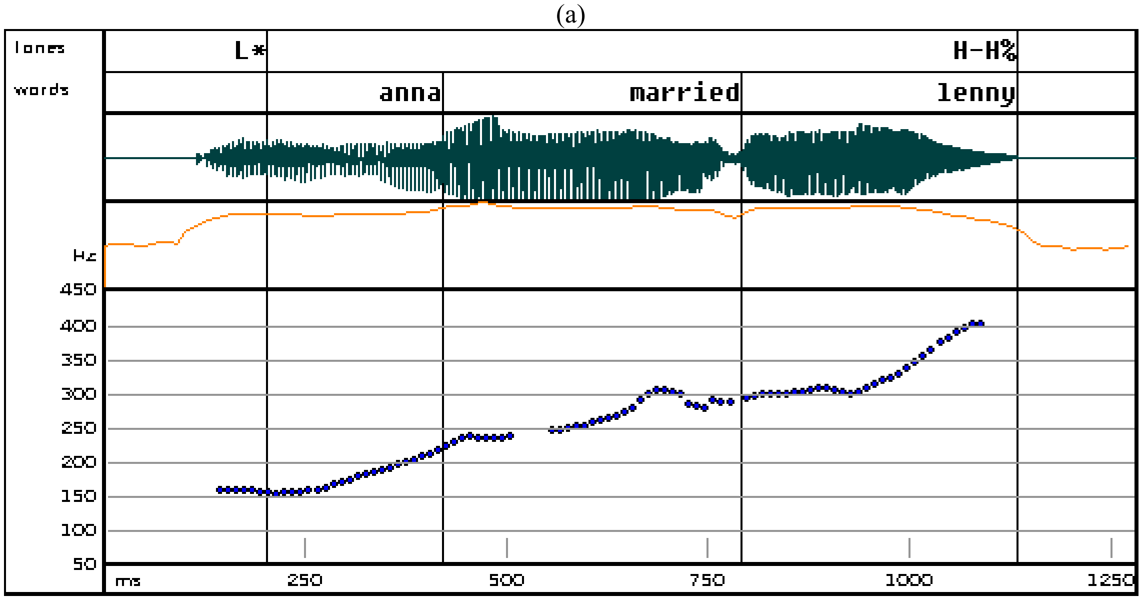
of the words that occur in these positions. Previous studies on voice quality and prosody will be discussed in depth in Chapter 2. Chapter 2 will also include a more detailed definition of voice quality. Chapter 3 will describe the methods used in this study and Chapter 4 will describe the data collection and analysis procedures. Chapters 5 and 6 will present and discuss the results.

Chapter 2: Background

Prosody uses continuous, phonetic variables to mark categorical differences in prominence and boundaries. For example, the primary correlate of intonation, one of the main components of prosody, is pitch. Pitch can be thought of as a continuous variable; it occupies a sliding scale from low to high and is linked to many paralinguistic phenomena, such as gender and emotion. Independent of any prosodic function pitch may have, it is known that pitch can be used to form phonological categories in lexical tone languages such as Mandarin (Sino-Tibetan) and Maasai (Nilo-Saharan). In languages without lexical tone, such as English, pitch is used to mark prosodic prominences and boundaries, as can be seen in Figure 2. The sentence in Figure 2a, a declarative, uses a high tone to mark the prominent word and a low tone to mark the sentence-final boundary. The sentence in Figure 2b, an interrogative, uses a low tone to mark the prominent word and high tone to mark the sentence-final boundary. Pierrehumbert & Hirschberg (1990) even propose that intonation tunes are compositional in meaning. For example, a high tone on a word conveys to the listener that information is new in the discourse; a low tone at the end of the phrase conveys to the listener a degree of finality.



(a)



(b)

Figure 2. The sentence *Anna married Lenny* produced with two different intonation tunes.

Voice quality shares many of these continuous and categorical characteristics with pitch. As noted above, voice quality can be thought of as occupying a continuum from breathy to creaky; it is also linked to paralinguistic phenomena, such as gender and emotion. And like pitch, there are numerous languages that use voice quality phonologically to make lexical contrasts, such as Zapotec (Oto-Manguean), Gujarati (Indo-European) and Hausa (Afro-Asiatic). Therefore, it is likely that voice quality is also used, along with pitch and duration, to mark prosodic boundaries and prominences.

2.1 Previous studies

Many previous studies have found evidence that segmental variation in speech is partially attributable to the prosodic location of the segments. These positions have been identified as prosodic domain-final, prosodic domain-initial and accented positions, and they can be marked by a large assortment of phonetic variables. For example, prosodic domain-final position is often marked by domain-final lengthening (e.g., Wightman et al. 1992). Prosodic domain-initial position can be marked with articulatory and acoustic strengthening (e.g., Fougeron & Keating 1997). Prominent positions can be marked with hyperarticulation (e.g., de Jong 1995). (See Cho (2001) and Fougeron (1999) for reviews.)

It is in precisely these positions that other studies have found voice quality variations in English not conditioned by particular segments. For example, creak is used to mark the ends of both paragraphs and sentences within paragraphs (Kreiman 1982; Lehiste 1975). Creak can even occur at the end of a single sentence in isolation (Henton & Bladon 1988). Voice quality changes have also been observed at the edges of smaller

prosodic units. Hagen (1997) found that creak occurs frequently either word- or phrase-finally. It has also been found that the rates of glottalization are significantly higher at full intonation phrase boundaries than at intermediate phrase boundaries (Redi & Shattuck-Hufnagel 2001). Dilley et al. (1996) found that allophonic vowel-initial-word creak occurred most frequently phrase-initially and when the word is accented. In other words, creak occurred most frequently at both the ends and beginnings of prosodic boundaries and when the word is accented.

This tendency for allophonic vowel-initial-word glottalization to occur more frequently in phrase-initial and accented positions can also be viewed as an instantiation of the articulatory strengthening commonly found in these positions. Strengthening is defined as an increase in articulatory effort and more consonant-like articulations for consonants and more vowel-like articulations for vowels (Pierrehumbert & Talkin 1992). Word-initial vowels may become glottalized or be preceded by a glottal stop (Dilley et al. 1996).

Speakers may also use a variety of voice qualities in these positions. Pierrehumbert & Talkin (1992) note that speakers can use either a breathy/soft or a laryngealized voice quality after the nuclear pitch accent in a sentence. Gobl (1988) found that speakers of Swedish use a breathy voice quality at the ends of sentences. Finally, it has been shown that utterance-final "creaky" voice is produced in a different manner (with a spread glottis and reduced lung pressure) than creaky voice produced elsewhere in normal English speech (*i.e.*, which is produced with a tightly adducted glottis) (Slifka 2000).

There are a number of limitations, though, to previous research on prosodic voice quality variations. Most of these studies only report on creakiness, and usually track creak by taking qualitative measurements of the time amplitude waveform, i.e. by labeling certain types of waveforms as creaky or subtypes of creaky (Dilley et al. 1996; Hagen 1997; Redi & Shattuck-Hufnagel 2001). The use of this technique will not capture other types of voice qualities that may be used in ways similar to more traditional creakiness, such as breathy voice or period doublings, or small, relative changes in voice quality. As a result, there will be an undersampling of phenomena.

Furthermore, these studies fail to track and control for all of the factors that could contribute to changes in voices quality. For instance, certain consonants in English have glottalized allophones and each speaker has their own individual voice qualities. These factors may contribute to the amount of creak, or any other voice quality, found in a corpus.

Moreover, most of the studies discussed above have focused on declaratives. Declaratives are most often characterized by high pitch accents and low boundary tones. As a result, it is difficult to tell from these studies if changes in voice quality are due to the presence of an accent or boundary or to the phonological pitch of that accent or boundary.

2.2 Defining voice quality for this study

Voice quality or phonation type has been defined as the laryngeal setting that co-occurs with the oral articulation of a vowel or consonant. In the literature, both the terms *voice quality* and *phonation type* are used to describe the vibration patterns of the vocal

folds, the acoustic results of these vibrations and listeners' perceptual evaluations of these acoustic patterns. The term *voice quality* has even been used to describe supralaryngeal settings that have distinct acoustic and perceptual results, such as *palatalized voice* or *pharyngealized voice* (Laver 1980). This study will use *voice quality* to cover only the acoustic and perceptual results of the state of the glottis. Furthermore, since the term *phonation type* has physiological overtones because of its association with the verb *phonate*, and no physiological measurements were made in this study, the term *voice quality* will be used.

It is often useful to regard different voice qualities as ranging along a continuum. Phonation types can range from the stiff vocal folds of laryngealized or creaky voice, to the "normal" vocal fold vibration patterns in modal phonation, to the relaxed and incomplete closure of the vocal folds in "breathy" or "weak" voice (Ladefoged 1971; Ladefoged 1973; Ladefoged & Maddieson 1996). Gerratt & Kreiman (2001), on the other hand, propose that some voice qualities are categorically distinct from modal voice while other voice qualities vary continuously from modal voice. For example, untrained listeners are able to successfully group together the period-doubled, from the amplitude-modulated and from the aperiodic voices in a rating task, suggesting that these voice qualities form perceptually distinct categories. Furthermore, these voice qualities are not only acoustically, but also physiologically distinct from each other and modal voice. Other voice qualities, such as breathy, in contrast, are not categorically distinct from modal voice. Expert listeners are unable to agree on how breathy a voice is and there are no clear acoustic or physiological correlates for breathiness. Finally, Gerratt & Kreiman

(2001) note that there is no clear definition for "modal" phonation either. Modal voice is often defined as the "default" voice quality, from which all other phonation types vary. They suggest that because some voice qualities vary categorically from "modal" and some voice qualities vary continuously from "modal" that this distinction between modal and non-modal voice quality may not be particularly useful.

Although there are not always clean physiological, acoustic or perceptual categories for the different voice qualities (Kreiman & Gerratt 1996;Gerratt & Kreiman 2001), it is still conceptually useful to regard voice qualities as occupying a continuum with breathy and creaky voice as the endpoints and modal voice quality being in the middle (Ladefoged 1971; Ladefoged 1973). This is because it is then possible to make relative comparisons among voice qualities on this continuum. In this study, the terms *creaky* and *breathy* will be reserved to describe the non-modal endpoints of the continuum, and those voice qualities that can be observed directly from waveforms or spectrogram. Within these non-modal endpoints, though, there is also a continuum of modal voice qualities, or glottal strictures, to use Ladefoged's (1971) term. This continuum depends on the tensions and compressions of the vocal folds. *Tense voice* is associated with high values of adductive tension, medial compression and longitudinal compression of the vocal folds; *lax voice* is associated with the opposite (Ladefoged 1971; Ní Chasaide & Gobl 1997). As noted above, this study is concerned with the range of voice quality variations that occur during all of speech, not just the "extreme" variations in voice quality that can be observed by qualitatively examining waveforms; therefore the terms

tense and *lax voice* will be used to describe and categorize this “middle” continuum of voice qualities.[1]

2.3 Topic of this study

The purpose of this study is to evaluate the effects of prosody on voice quality in American English. In light of the previous research reviewed above, the following four topics will be considered, using data from both interrogatives and declaratives:

1. **Correlation between voice quality and F0.**
2. **Effects of prominence on voice quality.**
3. **Effects of phrase edges on voice quality.**
4. **Effects of phonological tone on voice quality.**

2.3.1 Correlation between voice quality and F0

Several previous studies have suggested that there is a correlation between voice quality and/or vocal fold vibrations and phonetic pitch, since a high F0 is correlated with increased vocal fold tension and length and a low F0 is correlated with decreased vocal fold tension and length. For example, Ohala (1973) found that Hindi breathy-voiced stops are correlated with a lowered F0. Maddieson & Hess (1987) found significantly higher F0 for tense vowels in three phonation-type languages: Jingpho, Lahu and Yi.

However, while this correlation exists it is not necessary (Ladefoged 1973; Laver 1980). Ladefoged (1973) notes that changes in pitch are associated with contraction/relaxation of the cricothyroid muscle, whereas changes in phonation type are thought to be associated with rotation of the arytenoid cartilages. Thus, different voice qualities may occur on any phonetic pitch. For example, in Tianjin Mandarin, two tones (T1 and T3) are articulated in the lower pitch register, yet only T3 is associated with a

---

[1] The terms *tense* and *lax* have also been used to describe supralaryngeal vowel settings (c.f. Jakobson & Halle 1963). In this study, *tense* and *lax* will be used to describe laryngeal settings.

creaky voice (Davison 1991). There is even high speed film evidence of vocal fry occurring on a high F0 (Jody Kreiman p.c.). Also, Holmberg et al.'s (1989) study of pitch and voice quality in English did not find a strong correlation between any glottal parameters and F0. They did, however, find a strong correlation between some glottal parameters and sound pressure level, and sound pressure level and pitch. Titze (1994) also hypothesized that as the voice quality becomes either more tense or more lax the glottal source power (and thus the sound pressure level) decreases. In other words, moving very far away from modal voice quality in either direction decreases the glottal source power.

In summary, there is a strong tendency for high pitch and tense voice quality to be associated and low pitch and lax voice quality to be associated, although this tendency is by no means universal. This study therefore examines the hypothesis that voice quality is correlated with phonetic pitch in American English. In particular, it examines the hypothesis that there is a correlation between a high F0 and tense voice quality and low F0 and lax voice quality.

2.3.2 Effects of prominence on voice quality

As mentioned above, a number of previous studies have examined the effects of prominence on voice quality. For example, Gobl (1988), in his study of Swedish, found an increase in overall source spectral intensity (Excitation Energy) on focal syllables. Campbell (1995) found an increase in high frequency spectral energy for prominent words in English. Several other studies have examined the effects of prominence on allophonic word-initial glottalization. Pierrehumbert & Talkin (1992) found that

although stressed vowel-initial syllables had high rates of glottalization than their reduced counterparts, there was no effect of prosodic accentuation. On the other hand, both Pierrehumbert (1995) and Dilley *et al*. (1996), found increased levels of allophonic glottalization associated with prominence. It is likely that this kind of glottalization results in a tense voice quality.

This study therefore examines the effects of prominence on the voice quality of words and stressed syllables for all sentence intonations combined. It also examines the effects of prominence on words in declarative and interrogative sentences separately to see if there is a further effect of intonation tune. The following hypotheses will be tested:

- **Prominent words have a tenser voice quality than non-prominent words for all sentence intonations combined.**
- **Prominent stressed syllables have a tenser voice quality than non-prominent stressed syllables for all sentence intonations combined.**
- **Prominent words have a tenser voice quality than non-prominent words in declarative sentences.**
- **Prominent words have a tenser voice quality than non-prominent words in interrogative sentences.**

2.3.3 Effects of phrase edges on voice quality

As discussed above, several studies have examined the effects of phrase boundaries on voice quality. For example, Klatt & Klatt (1990) in an examination of declarative sentences found both increased tenseness (lower values of spectral tilt) and increased laxness (more noise in the region of F3) in utterance final position. Gobl (1988) found in Swedish a decrease in spectral intensity (EE) and an increase in dynamic leakage (RA) over the course of an utterance, indicating a laxer voice quality phrase-finally. Herman (1998) found that discourse-medial accented words have a higher amplitude than discourse-final accented words. Several studies have also examined the effects of

prosodic boundaries on the amount of allophonic vowel-initial glottalization and/or creaky waveforms. In general, they have found an increase in creak at phrase boundaries, both phrase-final and phrase-initial (Pierrehumbert & Talkin 1992; Pierrehumbert 1995; Dilley et al. 1996; Hagen 1997). Redi & Shattuck-Hufnagel (2001) found a hierarchical effect for phrase-final boundary creak, in that it occurred at higher rates at higher prosodic boundaries.

These studies have looked at both allophonic vowel-initial glottalizations (Pierrehumbert & Talkin 1992; Pierrehumbert 1995; Dilley et al. 1996) and syllable-medial creak (Hagen 1997; Redi & Shattuck-Hufnagel 2001) Although allophonic vowel-initial glottalization may be associated with a tenser voice quality, it is not clear what voice quality may be associated with creaky waveforms. Physiologically, period-doubled phonation is characterized by either tension or vibrational frequency asymmetries in the vocal folds (see Gerratt & Kreiman (2001) for a review). Vocal fry is thought to be produced with lax vocal folds that have a large contact area (see Baken & Orlikoff (2000) & and Laver (1980) for a review). Since studies of both modal and non-modal phonation have yielded a variety of results for the voice quality at phrase edges, it is not clear what should be expected from this study.

This study thus compares the effects of phrase edges on voice quality. There are a number of ways this effect can be studied: on a microscopic level, by comparing each syllable/word of a sentence; on a macroscopic level, by comparing large sections of a sentence with each other; or at a middle level, by comparing phrase-initial and phrase-final words with each other. As will be discussed later, the corpus for this study is

designed so that prominent words occur at the phrase edges. Therefore, this study will examine the effects of phrase edges on the voice quality of prominent words located at phrase-final and phrase-initial edges, both for all sentence types together and for interrogatives and declaratives separately. It will also examine the effects of phrase boundaries on large sections of the sentence. The following four hypotheses will be tested:

- **Phrase-initial prominent words have a different voice quality than phrase-final prominent words for all sentence intonations combined.**
- **Phrase initial sections of the sentence have a different voice quality than phrase-final sections of the sentence for all sentence intonations combined.**
- **Phrase-initial prominent words have a different voice quality than phrase-final prominent words for declarative sentences.**
- **Phrase-initial prominent words have a different voice quality than phrase-final prominent words for interrogative sentences.**

2.3.4 Effects of phonological tone on voice quality.

As discussed in the section on correlations, there is no a priori reason to suppose a relationship between pitch and voice quality. However, there is the possibility that a local phonological low or high, in other words a Low or High pitch accent, could cause a change in voice quality locally. For example, Hagen's (1997) study of English and German found that glottalizations are more likely to occur on words with a low F0. Furthermore, there is evidence in English that there is a positive correlation between subglottal pressure and pitch for boundary tones, in that subglottal pressure increases for High boundary tones and decreases for Low boundary tones (Herman et al. 1996). Thus, the following hypotheses will be tested:

- **Syllables with High pitch accents have a tenser voice quality than syllables with Low pitch accents.**
- **Words with High pitch accents have a tenser voice quality than words with Low pitch accents.**

- **Syllables with High boundary tones have a tenser voice quality than syllables with Low boundary tones.**

The next two chapters describe the methods used for testing the hypotheses outlined above. Chapter 3 explains the choice of methods used in this study and Chapter 4 will describe the data collection and analysis procedures. Chapters 5 and 6 will present and discuss the results.

Chapter 3: Methods

Differences in voice quality can be measured in a number of ways. Some of these are: perceptual assessment (e.g. Hirano 1981), qualitative assessment of the time amplitude waveform (e.g. Kirk et al. 1984; Dilley et al. 1996; Hagen 1997), qualitative assessment of the spectrogram (e.g. Kirk et al. 1984; Henton & Bladon 1988), quantitative measurement of the spectrum (e.g. Kirk et al. 1984, Ladefoged et al. 1988; Holmberg et al. 1995, Stevens & Hanson 1995; Hanson 1997; Hanson & Chuang 1999), and quantitative measurement of the voicing source (Javkin & Maddieson 1983; Huffman 1987; Ladefoged et al. 1988; Gobl & Ní Chasaide 1988; Löfqvist & McGowan 1992; Pierrehumbert & Talkin 1992). This chapter will review the selection of measurement techniques and present the methods for factoring out personal and segmental effects on voice quality. Chapter 4 will present the specific data collection and analysis procedures used in this study.

3.1 Perceptual vs. acoustic measurements of voice quality

Pitch, loudness and voice quality are all psycho-acoustic phenomena. Psychological phenomena, such as pitch, heaviness, sweetness, brightness, etc., can only be measured by asking people to make judgments, because they are interactions between people and objects, not properties of the objects themselves. In other words, the quality of heaviness is not inherent in a box but in the interaction between a box and the strength and experiences of a person lifting the box. Likewise, voice quality, is not an inherent property of a voice, but an interaction between a voice and a person listening to the voice.

Because voice quality is an interaction between a listener and a voice, ideally one would measure voice quality variations in sentences by asking expert listeners to give each syllable in the sentence a qualitative rating of voice quality, such as “breathy”, “rough”, “creaky”, etc. However, the qualitative description of voices assumes that the listeners’ perceptions can be associated with distinct acoustic components and that listeners are listening to each voice in the same way and are describing it using the same scale. Unfortunately, none of these assumptions is true. Kreiman & Gerratt (1996) found through a multi-dimensional scaling (MDS) analysis of 3160 similarity ratings of pathological voices made by twelve expert listeners, that the only stable dimension across all the listeners is the severity of the voice disorder (e.g. less severe vs. more severe). Among individual listeners, MDS gave a solution of clusters of voices, but no consistent voice quality could be found to characterize these clusters. Furthermore, Kreiman & Gerratt (2000) have found that the more complex a stimulus, the harder it is for listeners to agree on a category for it. They asked expert listeners to judge if a pathologic voice was primarily breathy or rough, and found that listeners could not agree on a category for a voice unless the voice belonged to a categorical extreme. They also found in a second experiment that expert listeners could not agree on the pitch of the pathological voices. In contrast, listeners were significantly better at agreeing on the pitch of synthetic stimuli that varied only in F0. These results indicate that listeners are unable to isolate and judge a single quality dimension, such as breathiness, roughness or pitch, while listening to a complex, natural voice. Finally, Kreiman & Gerratt (1998) found that when a pathological voice receives a moderate rating, this rating is not achieved by consensus

among raters, but as the average of the range of ratings assigned to this voice. These studies indicate that training expert listener ratings to evaluate voice quality differences in a sentence would not result in reliable, consistent measurements of voice quality.

3.2 Acoustic measurements of voice quality

Because voice quality is an interaction between a listener and a voice, aspiration noise or jitter does not measure voice quality any more than F0 measures pitch or amplitude measures loudness. However, there is a direct (but not linear) relationship between pitch and frequency, and loudness and amplitude. Even though the relationship between a particular voice quality and any acoustic measurement is much more complex and less well understood, it is still justifiable to use acoustic measurements to assess voice quality.

Although it is difficult to directly observe the phonation patterns of the glottis without resorting to invasive and expensive procedures (and thus take physiological measurements), it is possible to take acoustic measurements of the result of the vibrations of the vocal folds – the pattern of airflow through the glottis known as glottal flow. Figure 3 shows a schematic of vocal fold vibration and the resulting glottal flow. The amount of flow increases gradually as the folds open and then usually drops abruptly when the vocal folds close, since they close relatively abruptly. When the flow is at the zero line, the vocal folds are closed.

In the source-filter theory of speech production, the fundamental frequency and harmonics resulting from the vibration of the vocal folds are known as the voicing source. The frequencies of the harmonics in the spectrum are fixed by the fundamental frequency of the voice, but the amplitudes are not; it is these relative amplitudes of the

harmonics that give all characteristics of the voice beyond pitch (e.g. voice quality, vowel quality, etc.). The source may also have significant inharmonic (noise) components. The source is then modified by the acoustic filtering action of the vocal tract, whose precise nature depends on the vocal tract's shape and length. Thus, the vocal tract is often referred to as the *filter* (Fant 1960). In linear source-filter theory, it is assumed that the glottal source and the vocal tract do not interact.[2] The vocal tract is modeled as an all-pole filter shaping the source signal. There may also be anti-resonances or zeros for nasals or persistent glottal gaps. The radiation at the lips is modeled by a differentiator. The voicing source is then modified by the shape of the vocal tract to give us speech (see Figure 4).

In review, the amplitudes of the harmonics of a spectrum of a speech signal vary for two reasons – source variation and vocal tract filtering. In Figure 4, on the bottom left, there is an example of a voicing source with a spectrum with smoothly decreasing harmonics. Different voice qualities, however, will produce source spectra with different spectral tilts and shapes. The spectral tilt of a source spectrum will then carry over to some extent into the output spectrum of the speech signal (see Figure 4 bottom right). It is because both phonation type and the vocal tract filter affect amplitudes of harmonics in a spectrum of a speech signal that voice quality can be hard to decompose into these components.

[2] Although this is not entirely true (the source and filter do interact while the glottis is opening or if there is a persistent glottal gap), it is a useful and necessary assumption when attempting to separate the source and the filter.

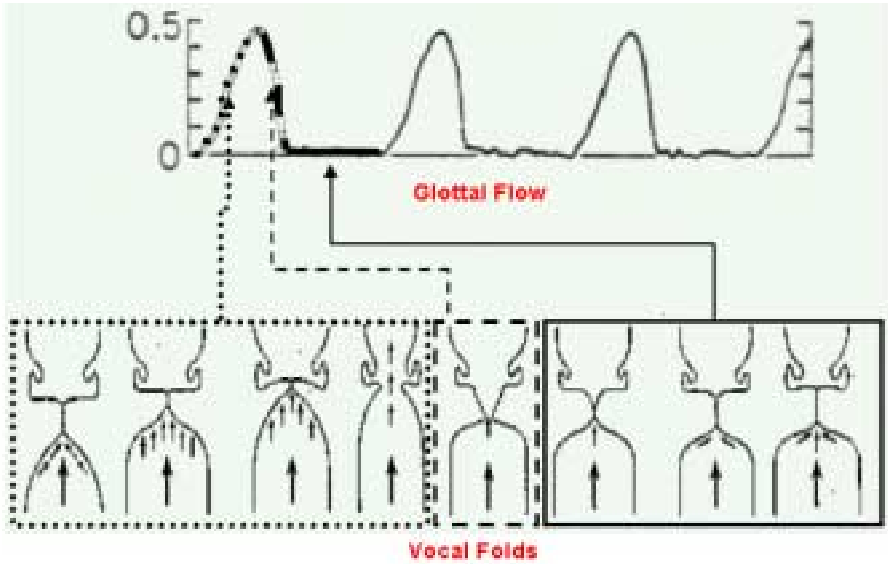


Figure 3. Schematic of vocal fold vibration and the resulting glottal flow. Time proceeds from left to right. Adapted from M. Rothenberg (1971), "The glottal volume velocity waveform during loose and tight glottal adjustments". In *The Proceedings of the VIIth International Congress of Phonetic Sciences*. Retrieved March 19, 2002 from http://www.rothenberg.org/Glottal.htm. And adapted from C.R. Schneiderman (1984), *Basic Anatomy and Physiology in Speech and Hearing*. San Diego, CA: College-Hill Press. p. 76. Retrieved March 19, 2002 from http://www.umanitoba.ca/faculties/arts/linguistics/russell/138/sec5/phonatio.htm.

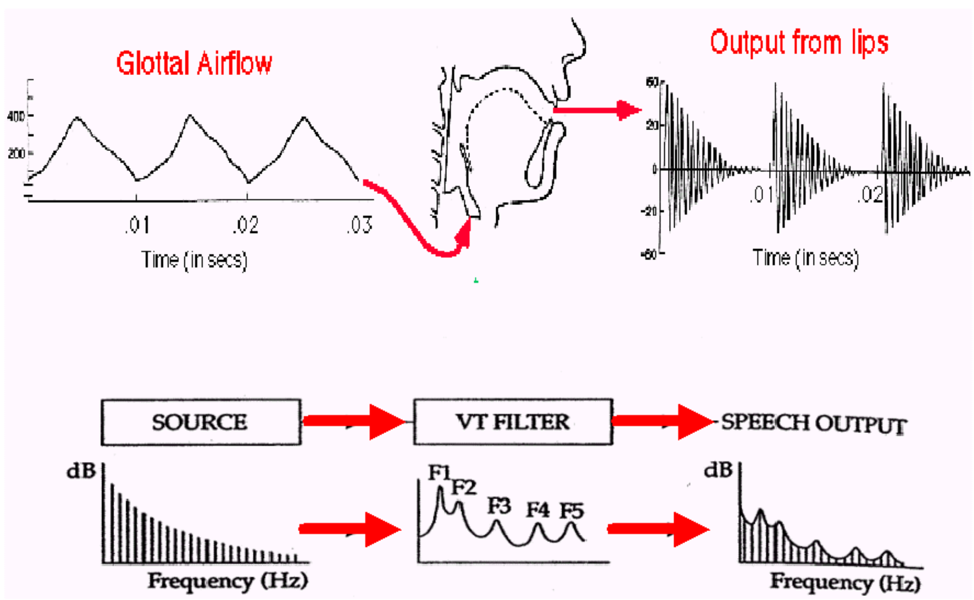
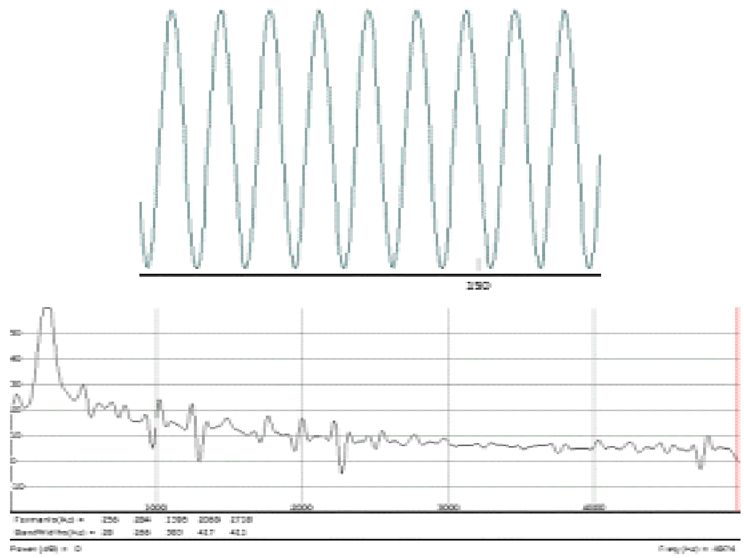


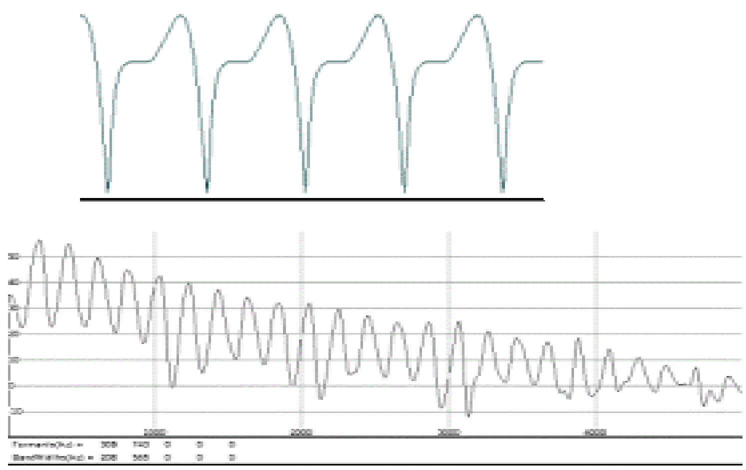
Figure 4. Schematic of the source-filter theory of speech production (Fant, 1960). Adapted from P. Lieberman, *Speech Physiology and Acoustic Phonetics*. New York: Macmillan. p. 32 and p. 35. Retrieved March 19, 2002 from http://www.ling.yale.edu:16080/ling120/Source_Filter/SFb.html. And adapted from A. Ní Chasaide and C. Gobl (1997), "Voice Source Variation". In W.J. Hardcastle and J. Laver, eds., *The Handbook of Phonetic Sciences*. Oxford: Blackwell. p. 430.

As noted above, though, there is no known direct relationship between any acoustic measurement and voice quality. If, however, an acoustic correlate of voice quality is changed incrementally in synthetic speech and corresponding changes occur in the perception of that speech, then it can be said that this acoustic measurement is "perceptually real". It has been shown in a number of studies that changing the shape of the glottal waveform in synthetic speech results in perceptually different voice qualities (Klatt & Klatt 1990; Childers & Ahn 1995; Gobl & Ní Chasaide 1999). Therefore, this study uses acoustic assessments of glottal flow as indices of changes in voice quality.

Voice source pulses for different voice qualities have different characteristic appearances. Gradual closure of the vocal folds, characteristic of a breathy or lax voice, causes the voice source pulse for a lax voice to be sinusoidal in appearance. The smoother and more sinusoidal the glottal pulse, the fewer higher harmonics there will be in the spectra of the source and speech output. Abrupt and complete closure of the vocal folds, characteristic of a tense voice, causes skewing and more abrupt changes in the shape of the glottal pulse for a tense voice. This is because the airflow through the glottis builds up gradually as the vocal folds open, but drops suddenly when the vocal folds close abruptly. The skewed shape of the pulse and the abrupt changes in its shape are correlated with an increase in the amplitude of the higher frequency harmonics in the spectra of the source and the speech output. (Bickley 1982; Childers & Ahn 1995; Hanson 1997; Huffman 1987; Ladefoged 1983; Ní Chasaide & Gobl 1997). See Figure 5 for details.



(a)



(b)

Figure 5. Differentiated glottal pulse shapes and their associated spectra. (a) shows a sinusoidal pulse with a spectrum dominated by a single, low frequency harmonic. (b) shows a pulse with a more complex shape with high frequency harmonics.

3.2.1 Methods for assessing the shape of glottal flow

There are two ways of quantitatively assessing the shape of the glottal flow using acoustic measurements. One is through comparison of the amplitude of different harmonics in the spectrum of the speech output (e.g. Kirk et al. 1984; Ladefoged et al. 1988; Holmberg et al. 1995; Stevens & Hanson 1995; Hanson & Chuang 1999; Hanson

1997). The second is through quantitative measurement of the voicing source (Javkin & Maddieson 1983; Huffman 1987; Ladefoged et al. 1988; Gobl & Ní Chasaide 1988;Löfqvist & McGowan 1992; Pierrehumbert & Talkin 1992).

Spectral measurement of the speech signal compares the amplitudes of different harmonics in the spectrum of a sample of speech output. Typical spectral measurements include comparisons between the first and second harmonics (H1-H2) and comparisons between the first harmonic and the strongest harmonic in the third formant (H1-A3). For breathy voices the first harmonic usually has the highest amplitude; for creaky voices the higher frequency harmonics have higher amplitudes. The primary advantage of spectral measurement of the speech signal is they may be calculated from any recording made with a good quality microphone, and most acoustic software is equipped to make Fourier spectra of the acoustic signal. Furthermore, the relationship between such measures and the underlying glottal pulse characteristics is fairly well established (e.g. Hanson 1997). On the other hand, there are two drawbacks to spectral measurements. First, they are sensitive to differences in F0. For the same vowel spoken on different pitches, the relationship between the second harmonic and the first formant can change, causing the amplitude of this harmonic to be amplified or attenuated. For example, the value for H1-H2 for the spectrum in Figure 6 is 10 units and the value for H1-H2 for the spectrum in Figure 7 is 5 units. These differences in H1-H2 for the two spectra are not due to a difference in voice quality, but to a difference in F0.

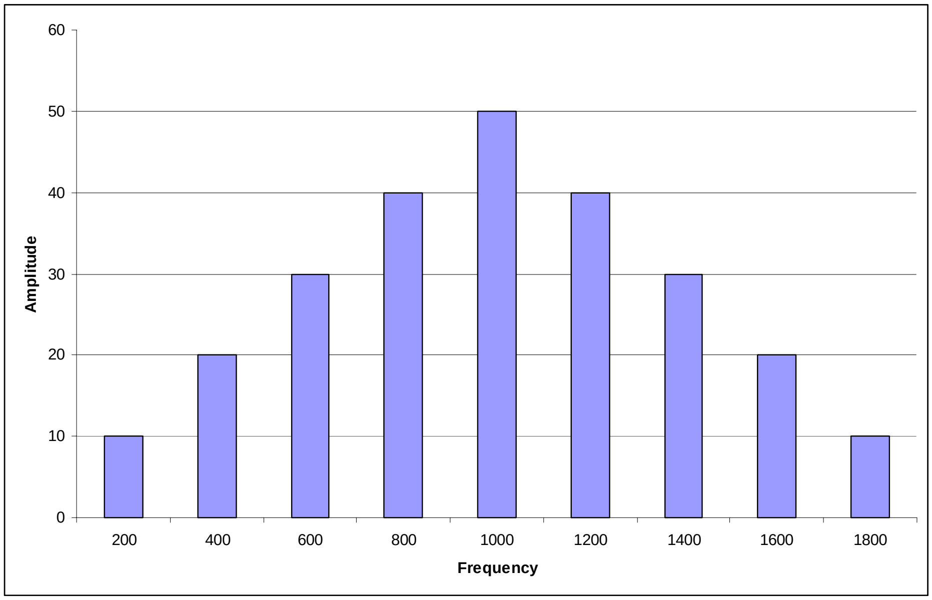


Figure 6. An idealized spectrum of a formant with an F0 at 200 Hertz. Note that the amplitude difference between the first and second harmonics is 10 units.

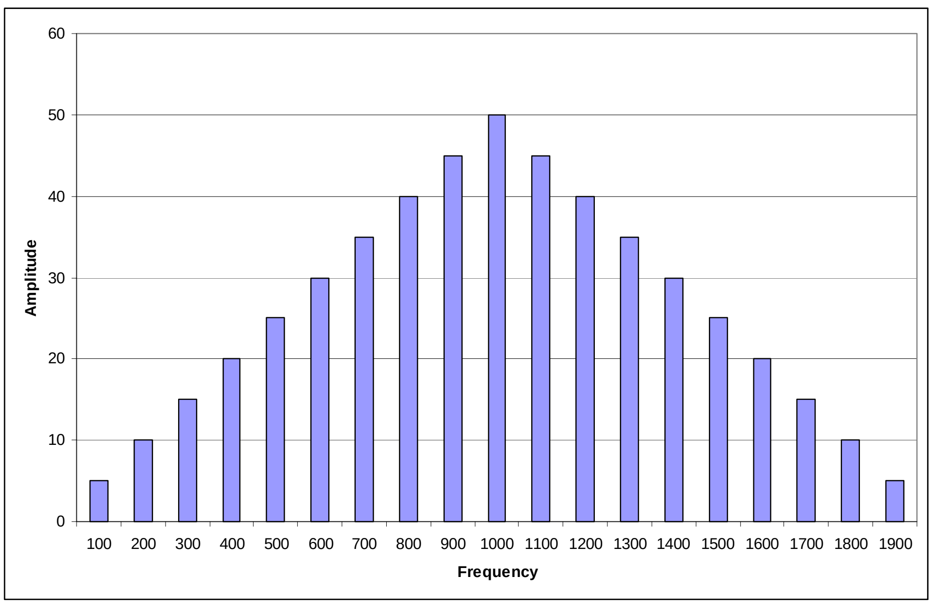
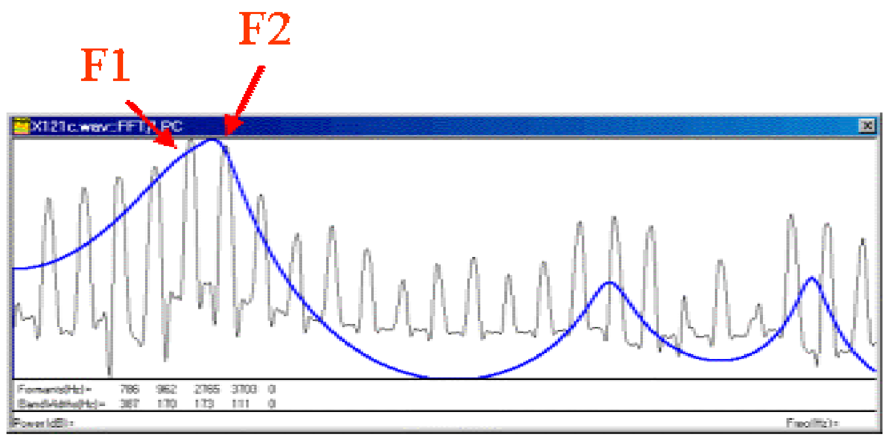
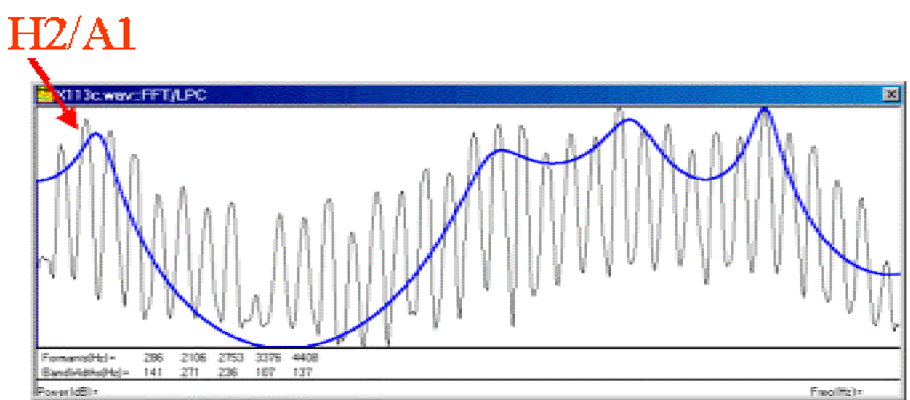


Figure 7. An idealized spectrum of a formant with an F0 of 100 Hertz. Note that the amplitude difference between the first and second harmonics is 5 units.



(a)



(b)

Figure 8. Examples of spectra showing the interaction between formant frequency and formant amplitude. (a) shows a spectrum of the vowel /ɑ/ with very close first and second formants. (b) shows a spectrum of the vowel /i/ with the second harmonic being the strongest harmonic in the first formant.

Second, these spectral measurements are sensitive to differences in vowel qualities. Vowels with close formants may amplify harmonics in the region of the two formants or make it difficult to identify harmonics associated with particular formants. For example, the vowel /ɑ/ has F1 & F2 close together and the vowel /i/ has F2 & F3 close together. Low first formants may also amplify the first or second harmonic. See Figure 8 for examples. These effects may even be so strong that different vowels appear to have different voice qualities (Epstein & Payri 2001).

As an alternative to spectral measurements, one can evaluate the glottal flow by mathematically inverting and removing the vocal tract filter from the speech output. This process is called inverse filtering (see Figure 9). There are two ways to collect speech for inverse filtering. One is to collect oral airflow with a face mask. Inverse filtering oral airflow results in true glottal flow (see Figure 10). Using a mask, however, distorts frequencies over 3500 Hz, which would result in the loss of high frequency spectral noise information and high frequency source information. Alternatively, one can collect measures of speech pressure using a high quality condenser microphone. Inverse filtering speech pressure results in differentiated glottal flow because of the radiation characteristic of the lips. The differentiated glottal flow is thus the first derivative of the true glottal flow. In practice, speech is usually collected with a microphone and then inverse filtered. One can then work directly with the differentiated glottal flow or integrate the differentiated glottal flow to obtain the true glottal flow.

The primary limitation of inverse filtering is that it requires accurate vocal tract filter estimations. Fant (1995) notes a case when two different calculations of the filter resulted in source pulses of different shapes. Vocal tract estimation may be further complicated by significant noise components or source-vocal tract interactions in the signal. Because of these interactions, the output of the inverse filtering procedure often contains theoretically undesirable bumps and ripples. To resolve these difficulties, the output of the inverse filtering procedure is usually fit with a theoretical model of the glottal pulse, and measurements of the source are usually taken from the model and not

the inverse filter output (see Figure 12). In this study the Liljencrants/Fant (LF) model of differentiated glottal flow will be used.

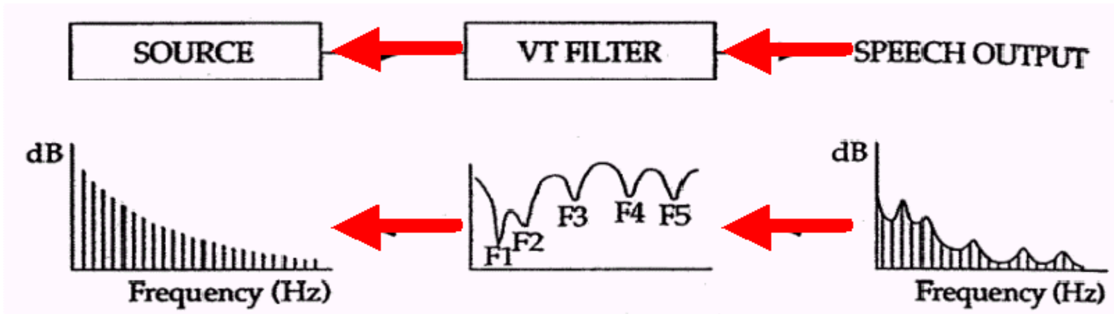


Figure 9. Schematic of the inverse filtering procedure. The inverse filtering process occurs from right to left. Adapted from A. Ní Chasaide and C. Gobl (1997), "Voice Source Variation". In W.J. Hardcastle and J. Laver, eds., *The Handbook of Phonetic Sciences*. Oxford: Blackwell. p. 430.

Figure 10. Schematic of inverse filtering of oral airflow collected with a face mask. Adapted from A. Ní Chasaide and C. Gobl (1997), "Voice Source Variation". In W.J. Hardcastle and J. Laver, eds., *The Handbook of Phonetic Sciences*. Oxford: Blackwell. p. 430.

Figure 11. Schematic of inverse filtering of speech pressure collected with a condenser microphone. Adapted from A. Ní Chasaide and C. Gobl (1997), "Voice Source Variation". In W.J. Hardcastle and J. Laver, eds., *The Handbook of Phonetic Sciences*. Oxford: Blackwell. p. 430.

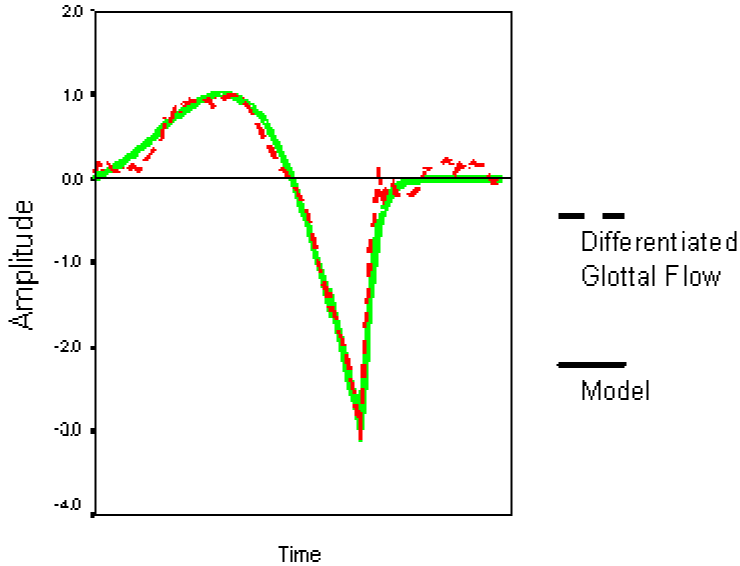


Figure 12. Sample of differentiated glottal flow and a theoretical model fit to it.

3.2.2 The LF model of the glottal flow derivative

The LF model (named for its creators Johan Liljencrants and Gunnar Fant) is a mathematical model of the glottal flow derivative characterized by a combination of a sinusoid and an exponential (Fant et al. 1985). Modeling the glottal flow derivative instead of the true glottal pulse has two advantages. First, it could involve introduction of further error through simplifying assumptions involved in the implementation of the integration of the output of the inverse filtering procedure. Second, since the glottal flow derivative is a derivative, it is easier to mark the changes in the shape of the glottal flow and thus determine points of interest on the glottal flow, such as the point of maximum closing speed (see point D in Figure 13).

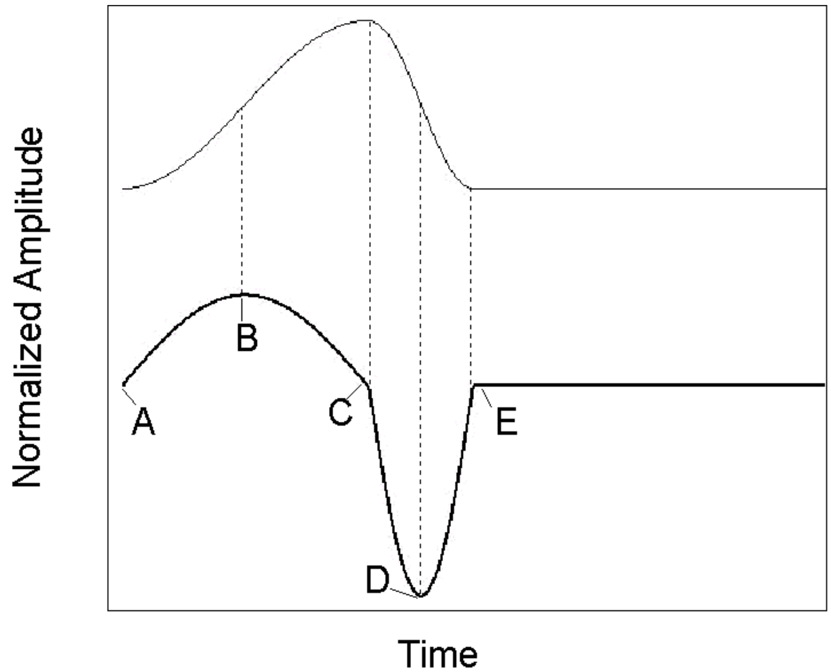


Figure 13. Schematic of models of true glottal flow (top) and differentiated glottal flow (bottom). Opening phase is from A to C. Closing phase is from C to E. Note how the differentiated glottal flow marks points of maximum discontinuity in the opening phase and closing phase with a maximum (B) and minimum (D). The segment from D to E on the differentiated glottal flow is called the return phase.

In calculating the LF model, several points are chosen on the glottal flow derivative (U′), which become the parameters for minimizing error between the LF model and the differentiated glottal flow. It is important to note that it is the mathematical equations used to model the source that characterize the LF model. The parameters chosen to calculate these equations are often estimations of major features of the glottal source, but their selection is inherently arbitrary – and other parameters can be chosen. Furthermore, the parameters that are chosen to calculate the LF model are theoretical, and do not correspond directly to any physiological states. However, it is common practice to relate these parameters to states of the glottis or the glottal flow, such as open phase, glottal

skew, etc. The points used in calculation of the LF model are usually the following major features of the glottal flow derivative (U′) (see Figure 14):

t_0 = the start of the pulse
t_c = the length of the entire pulse
t_p = the length of the time that U′>0
t_e = the time of the maximum negative value of U′
E_e = the value of the maximum negative U′
t_a = the effective duration of the return phase (the segment from D to E in Figure 12); this duration is determined by the projection onto the time axis of a tangent to the exponential curve at time t_e

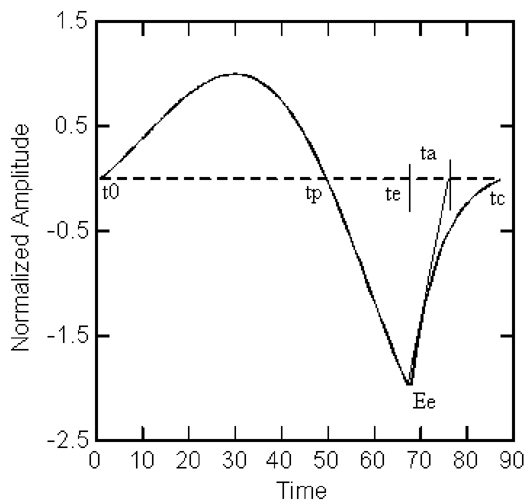


Figure 14. Schematic of traditional LF model of differentiated glottal flow.

In this study, the LF model used differs from the traditional LF model. It has been modified, particularly in the modeling of the return phase, in order to better fit non-modal phonation (see Figure 15). Ta, the duration of the return phase, has been changed to t2, the time increment to 50% decay in the return phase (i.e. where a line intersecting the halfway mark on the return phase intersects the time axis). The traditional LF model also has an equal area constraint – the area under the positive half of the curve must be equal to the area under the negative half of the curve. For many non-modal voices

implementing the equal area constraint resulted in an LF pulse with a return phase that returned sharply to zero (see Figure 16), which introduced high frequency artifacts into the signal. See Kreiman et al. (2000) for a detailed review of the traditional and revised versions of the LF model.[3] Calculation of the LF model will be discussed in-depth in Chapter 4.

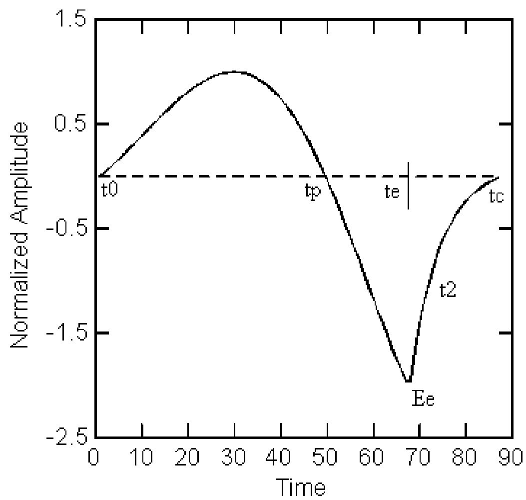


Figure 15. Revised LF model of differentiated glottal flow. Note that t_2 has replaced t_a.

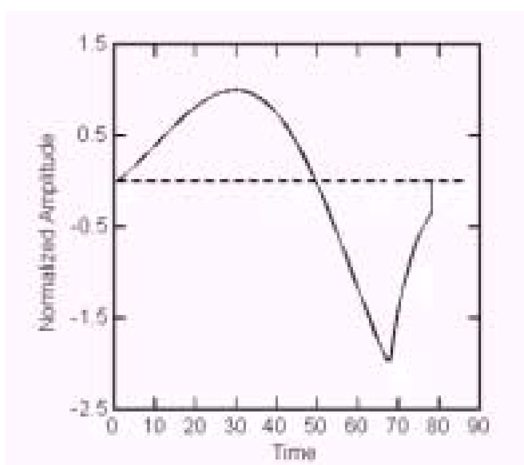


Figure 16. Example of an LF pulse with a return phase that returns sharply to zero.

[3] Software for inverse-filtering and LF-fitting is available from the Bureau of Glottal Affairs on the following website: http://www.surgery.medsch.ucla.edu/glottalaffairs/software_of_the_boga.htm or by contacting Dr. Jody Kreiman at jkreiman@ucla.edu.

3.2.3 Assessing the LF model of the glottal flow derivative

The shape of the glottal pulse is further qualified using the following traditional measurements (adapted from Ní Chasaide & Gobl (1997), Gobl & Ní Chasaide (1992) and Gobl & Ní Chasaide (1988)):

- **Open Quotient, OQ** OQ is measured as the ratio of the opening phase of the glottal waveform to the period of the pulse, or $\frac{(te-t0)}{(tc-t0)}$. This measurement predicts the values for the amplitudes of the lower harmonics; an increased value of OQ is correlated with an increase in amplitude of the lower harmonics in the voice spectrum.[4]

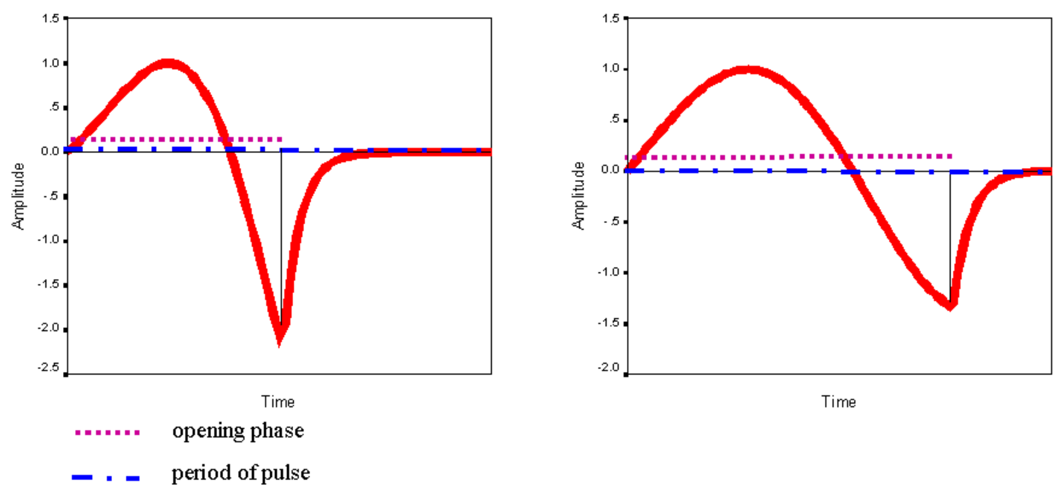


Figure 17. Schematic of open quotient (OQ). Open quotient is the ratio of the opening phase to the period of the pulse. The figure on the left shows a glottal pulse with a small value for OQ, the figure on the right shows a glottal pulse with a large value for OQ.

[4] Open quotient may alternatively be calculated as the ratio of the *open* (not opening) phase to the period of the pulse. The open phase is a measure of the entire time the vocal folds are open.

- **Glottal Symmetry/Skew, RK** Glottal skew measures the relationship between the closing phase and opening phase of the differentiated glottal pulse. It is measured as the following ratio (see Figure 18): $\frac{((te-t0)-(tp-t0))}{(tp-t0)}$. It is the inverse of the speed quotient. The glottal pulse is normally skewed to the left because the opening phase is longer than the closing phase. Acoustically, the skewing of the pulse affects the amplitude of the low frequency harmonics, and the more symmetrical the pulse, the greater the amplitude of the low frequency harmonics. Also, a more symmetrical pulse tends to have a more pronounced weakening of certain harmonics (spectral notches). Ní Chasaide & Gobl (1997), however, claim that this is probably due to an associated decrease in excitation strength rather than to the glottal skew.

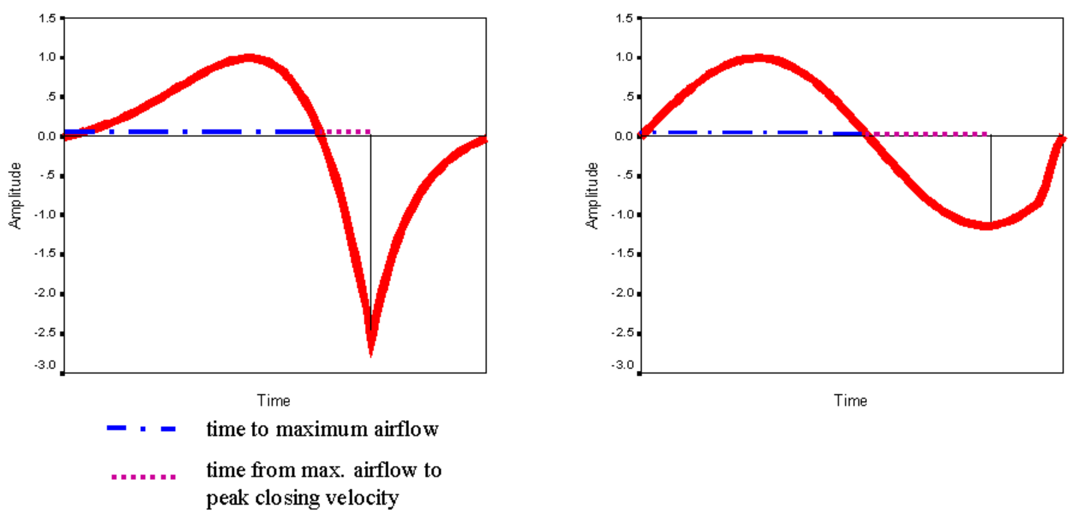


Figure 18. Schematic of glottal symmetry/skew (RK). Glottal symmetry is the ratio of the closing phase to the opening phase of the pulse. The figure on the left shows a glottal pulse with a small value for RK, the figure on the right shows a glottal pulse with a large value for RK.

- **Spectral Intensity/Excitation Strength, EE** EE is measured as the amplitude of the negative peak of the differentiated glottal pulse. This peak occurs at the maximum closing velocity of the true glottal pulse (Figure 13). Acoustically, EE is correlated with the overall intensity of the signal.

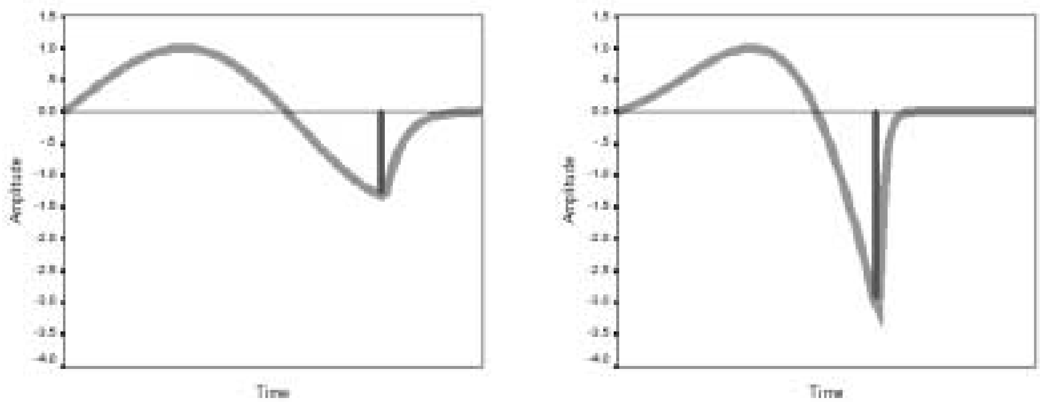
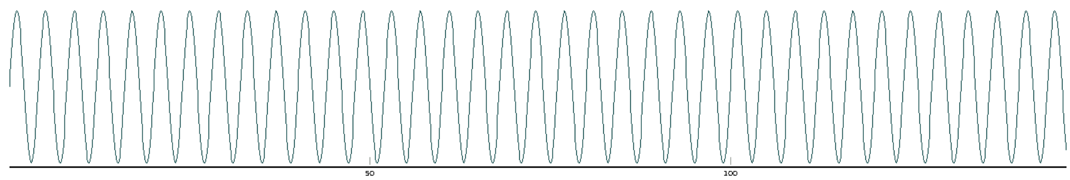


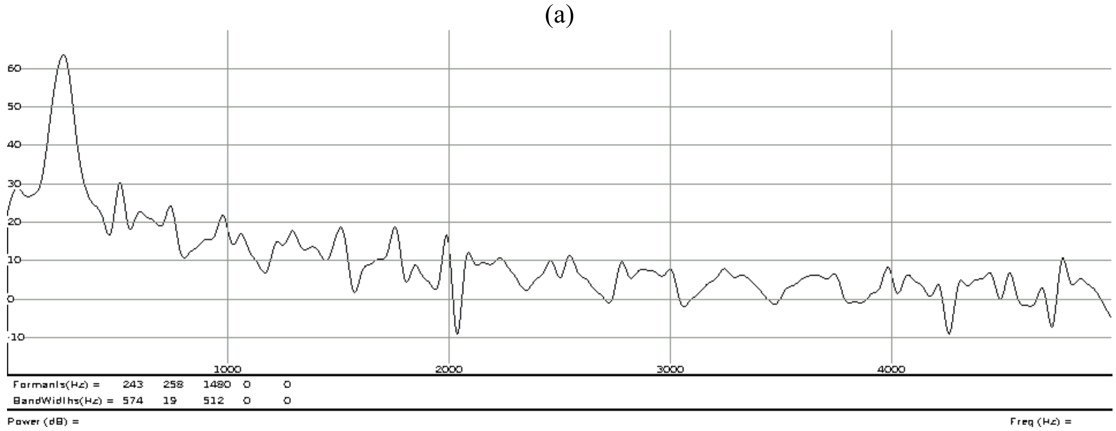
Figure 19. Schematic of excitation strength/spectral intensity (EE). Excitation strength is the amplitude of the negative peak of the differentiated glottal pulse. The figure on the left shows a glottal pulse a small value for EE, the figure on the right shows a glottal pulse with a large value for EE.

A fourth measurement, Dynamic Leakage or RA, is often taken as well. RA is measured as the ratio of the time constant of the return phase, ta, to the period of the pulse. At the production level, this measurement is related to the manner in which the vocal folds make contact during phonation, i.e. instantaneously or in a gradual fashion. Acoustically, differences in dynamic leakage affect the spectral slope of the signal. In the LF model used in this study, however, ta has been replaced by t2, thus making it difficult to directly measure RA from the parameters used in the model. Therefore, RA is replaced by a direct measurement of the spectral slope of the differentiated glottal flow, “Spectral Linearity” (Linearity). Spectral linearity is calculated in the following manner. A line is created by connecting the peaks (dB) across the entire FFT spectrum of a glottal

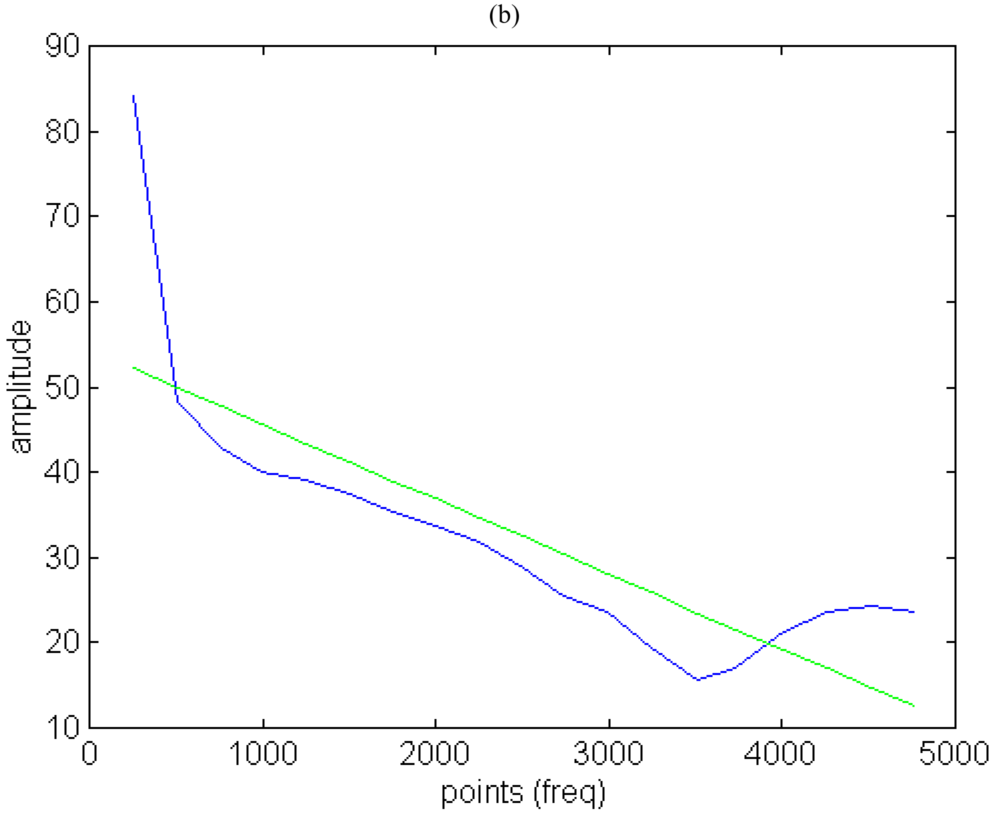
pulse. The value of the squared correlation, r^2, for the regression is then calculated. Regression analysis measures how close the line fit to the spectrum is to a straight line; a straight line implies that the glottal spectrum decreases gradually without any strongly attenuated frequencies. The value of spectral linearity is expected to be small for sinusoidal pulses, because they have very strong low frequency components, but very weak high frequency components (see Figure 20). The value of spectral linearity is expected to be large for glottal pulses that are skewed and/or have a closed phase, because these features amplify high frequencies in the spectrum (see Figure 21). Spectral linearity was calculated for differentiated glottal flow for Epstein's (1999) study comparing linguistic and pathological breathiness. Like other LF measurements used in that study, spectral linearity detected a difference between pathologically breathy tokens and linguistically modal tokens, but not between pathologically breathy tokens and linguistically breathy tokens.   As expected, modal tokens had a larger value for spectral linearity than breathy tokens. Also, in that data set, spectral linearity was correlated with EE and RK, but not OQ.



(a)

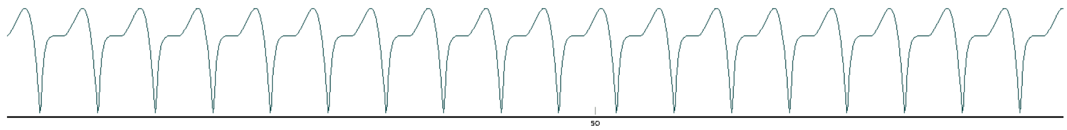


(b)

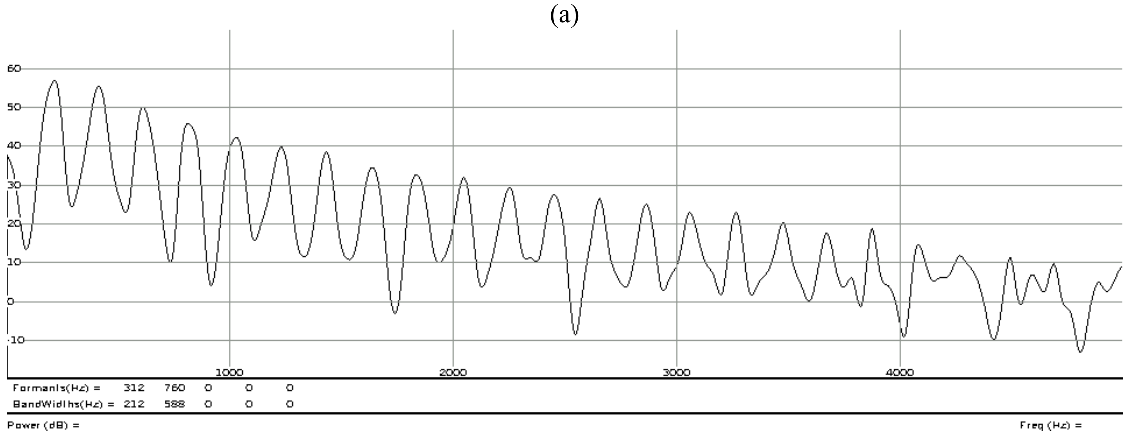


(c)

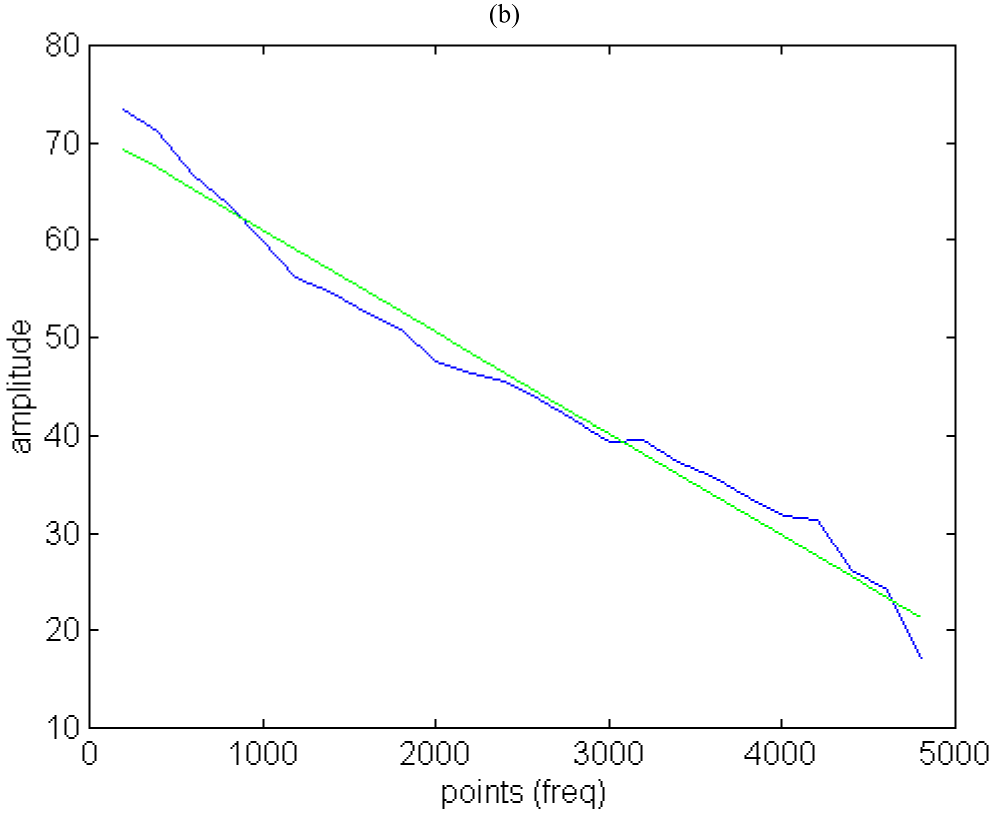
Figure 20. Differentiated glottal flow (a), dB FFT spectrum (b) and line fit to the spectrum (c) for a glottal pulse with a low value of Linearity at 0.64.



(a)



(b)



(c)

Figure 21. Differentiated glottal flow (a), dB FFT spectrum (b) and line fit to the spectrum (c) for a glottal pulse with a high value of Linearity at 0.98.

It is important to note that all four measurements used in this study are relative measurements to control for differences in glottal flow derivative amplitude and length. OQ and RK are ratios; EE is measured after the positive peak of the LF model of the glottal flow derivative has been normalized to a value of one; the value of r^2 for any regression analysis is always between zero and one.

In summary, lax voice is characterized by loosely vibrating vocal folds that produce smooth and sinusoidally shaped glottal flow. It is expected that lax voices will have large values for open quotient (OQ) and glottal symmetry (RK), and a smaller values for spectral intensity (EE) and spectral linearity (Linearity). Tense voice is characterized by greater vocal fold tension and skewed glottal pulses. It is expected that tense voices will have smaller values for open quotient and glottal symmetry, and larger values for spectral intensity and spectral linearity (Ní Chasaide & Gobl 1997). See Figure 22 for a visual representation.

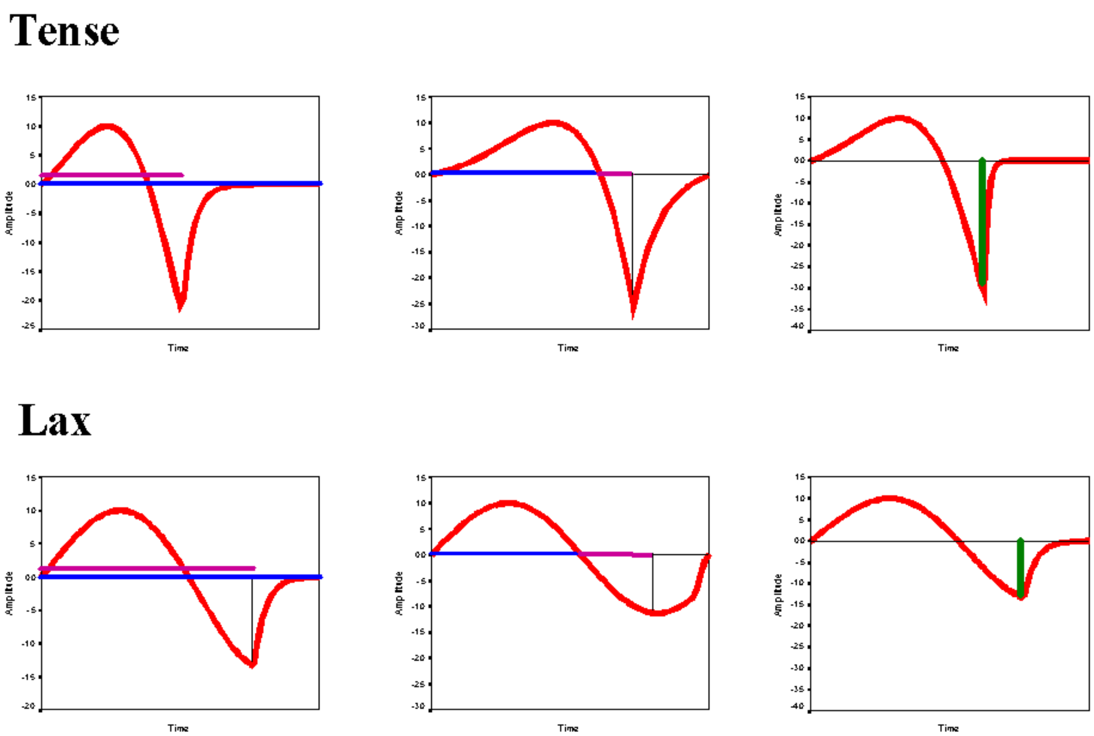


Figure 22. Comparison of LF measurements between tense and lax voices.

3.3 Factoring out linguistic and personal voice quality variations

It is possible that each vowel possesses its own intrinsic voice quality. After all, the muscles that lower the mandible – the anterior belly of the digastric, the geniohyoid and the mylohyoid – are attached on their inferior end to the hyoid bone, which is in turn attached to the thyroid cartilage of the larynx via the thyrohyoid muscle (The UCLA Phonetics Laboratory 1990). Thus, movement of the jaw can affect the position of the larynx and tension in the vocal folds. In other words, vowel height and frontness could potentially affect the glottal setting of the vowel. However, the pitch changes that are caused by the effects of jaw position on vocal fold tension are quite small, only on the order of a few Hz, implying that the changes in vocal fold tension are quite small, as well (Lehiste 1970; Vilkman et al. 1991). This suggests that there would be only minimal, if any, voice quality changes associated with a change in jaw position.

Although intrinsic vocalic voice quality differences may be minimal, several studies have, on the other hand, found measurable effects of consonants (or lack of consonants) on the voice quality of their surrounding vowels. Although the source is generally stable for a vowel, there are changes in the source in the transitions to and from consonants. These changes are always greater in transitions to and from voiceless consonants than voiced consonants (Löfqvist & McGowan 1992). Vowels in English are also more likely to be breathy before voiceless stops and after /h/ (Gobl & Ní Chasaide 1988; Epstein 1999). Furthermore, both allophonic and phonemic voice qualities do not always extend through the entire vowel (Blankenship 1997; Epstein 1999).

Along these same lines, there are significant voice quality variations from person to person. For example, several studies have found that women's speech is breathier than men's speech (Henton & Bladon 1985; Klatt & Klatt 1990; Hanson & Chuang 1999). Hanson & Chuang (1999) found that several of their male subjects had period doubling unlike the female subjects in Hanson's earlier studies (Hanson 1997). Even the tendency to use breathy and laryngealized allophones varies from person to person. Ní Chasaide & Gobl (1988) found that some native speakers of English had breathy vowels before voiceless consonants whereas other speakers had breathy vowels for only a few glottal pulses before voiceless consonants. Dilley et al. (1996) found that speakers' rates of glottalization of word-initial vowels vary from 13% to 44%. Redi and Shattuck-Hufnagel (2001) in their analysis of glottalizations at prosodic boundaries, found that glottalization rates vary from 13% to 88% in one corpus and 51% to 80% in another corpus. Byrd (1992) found that there were both gender and regional differences in the production of allophonic glottal stops among native speakers of American English.

The results of the studies discussed above suggest that it is currently impossible to disentangle prosodically varying changes in voice quality from changes in voice quality caused by influences of surrounding segments or underlying personal voice quality. Comparisons of voice quality measurements across words within a single speaker would be affected by the vocal tract filter of each vowel and the surrounding consonants. The voice quality measurements for the vowels in "had", "who'd", "bad" and "bat" could show voice quality differences even without variations in prosodic position. Comparisons of any single word (never mind multiple words) across speakers may be

influenced by each person's underlying voice quality and their tendencies towards certain allophonic segmental variations. One person's end-of-sentence creak may be another person's normal voice quality, and yet another person's allophone of coda /t/.

3.3.1 The test and base corpora

To factor out the interactions between personal/segmental voice quality differences and prosodic voice quality differences, the corpus in this study contains two types of sentences. The first type are "test" sentences, all composed of the same sequence of words. These sentences are either declaratives or interrogatives and they vary the location of the accented word in the sentence. The test corpus consists of the following sentences, the bold word being accented and receiving narrow focus:

- **Dagada** gave Bobby doodads.
- Dagada gave Bobby **doodads**.
- **Dagada** gave Bobby doodads?
- Dagada gave Bobby **doodads**?

Data from these sentences will help identify if a change in voice quality is due to the overall intonation tune of the sentence or to the location of the accented word in the sentence.

The sentence *Dagada gave Bobby doodads* is designed so all vowels are surrounded by voiced consonants. This is done to avoid breathy vowels caused by aspiration, glottalized word-initial vowels and glottalized vowels before glottal allophones of voiceless coda stops. There are also no nasals (avoids nasal zeros in the spectrum) or approximants (which are often syllabic or difficult to separate from the vowel). Most of the consonants are stops to ease the distinction between consecutive consonants and

vowel in the waveform and spectrograms. Furthermore, the sentence, *Dagada gave Bobby doodads*, begins and ends with unstressed syllables to assist in disentangling word-level or sentence-level accent effects from phrase edge effects on voice quality. Finally, the sentence does not contain any potentially reduceable function words.

The second type are "base" sentences. The purpose of these sentences is to give a measurement of each speaker's general pronunciation tendencies for each word in the test corpus. Since the test corpus consists of variations of the sentence *Dagada gave Bobby doodads*, the base corpus consists of the following sentences:

- **Mary** said "Dagada" today.
- **Nancy** said "Bobby" today.
- **Dana** said "gave" today.
- **Peter** said "do-dads" today.

These sentences are designed so that each base sentence and each base word are said as identically as possible. Speakers are asked to accent the bolded name (*Mary*, *Dana*, etc.). As a result, the word of interest (*Dagada*, *gave*, etc.) has a weaker accent and does not have an extreme pitch excursion (see Figure 23).

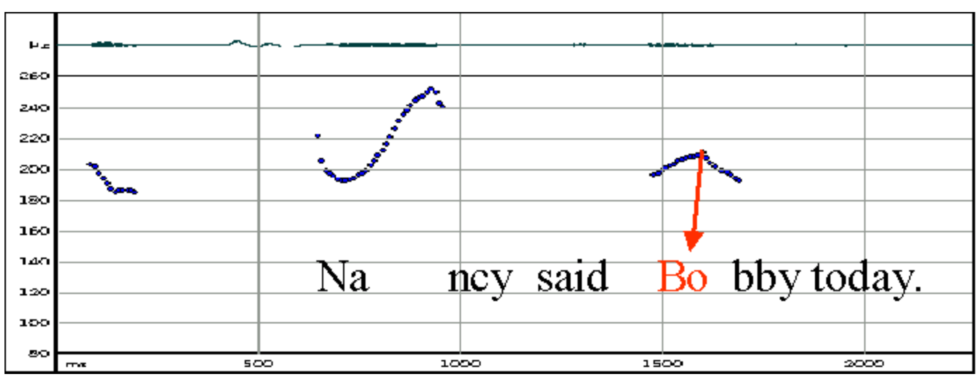


Figure 23. Pitch track from a base sentence.

To factor out the segmental and personal voice quality differences of the test sentence *Dagada gave Bobby doodads*, each word in the test sentence is compared to itself as pronounced in the base sentence. The measurements from these words are then changed into normalized measurements. In the normalized measurements, the value of the difference between the test and base words is divided that by the value of base word: $\frac{\text{test} - \text{base}}{\text{base}}$. Normalized measurements are percentages of the baseline values of each word in the corpus. A negative normalized measurement implies that the test word has a smaller measurement than the base word. A positive normalized measurement implies that the test word has a larger measurement than the base word. It is important to note that there is no claim being made for any special status of the voice quality of the baseline values – they are only providing a reference value against which the test values can be compared.

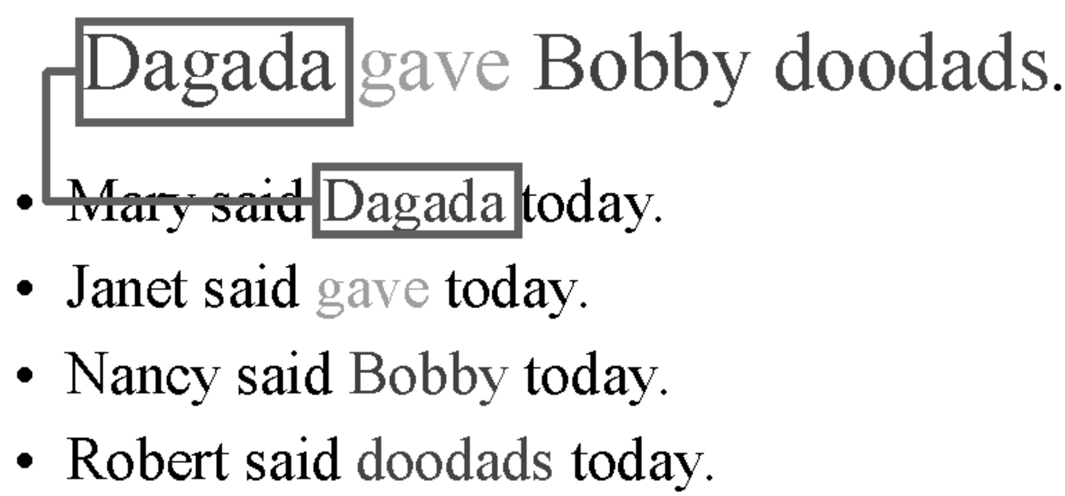


Figure 24. Comparing the test and base words.

Chapter 4: Procedures

This chapter describes the recording and analysis procedures used in this study. Results for a pilot study of these procedures are reported in Epstein (2001).

4.1 Subjects

Three subjects were analyzed for this study, two women (B and S) and one man (L). The subjects were between 25 and 35 years of age and all were native speakers of Western American English. Each was compensated $10 for participating in the study.

4.2 Recording procedures

Signals were transduced with a high-quality 1.0” Bruel & Kjaer condenser microphone placed 5 cm from the subjects’ lips. The B&K microphone was used because it has a very low frequency response (to DC), thus preserving the phase information of the signal necessary for calculating the shape of the glottal pulse. Subjects were recorded in a sound booth. Signals were sampled at 20 kHz and downsampled to 10 kHz. All data from a subject were collected during a single session.

Subjects were asked to speak at a comfortable loudness level, and were asked to repeat sentences that did not have the proper intonation pattern. Base sentences were recorded first. The base sentences were interspersed with selections from the Harvard sentences (a list of sentences developed to test speech intelligibility in noise (IEEE 1969)). The Harvard sentences also had a word boldfaced for narrow focus. See Table 1 for an example. These extra sentences were inserted to encourage resetting of the subjects pitch range in between recitations of the near identical base sentences.

Test sentences were then recorded in 10 blocks of six sentences, with the order of the six sentences randomized within each block. Each test sentence was preceded by a short scenario to encourage the proper intonation pattern. Following each block of test sentences, subjects read aloud one or more short poems to encourage resetting of their pitch ranges. See Table 2 for an example of the scenarios presented with the test sentences.

Sentence type	Sentence
Harvard sentence	The **birch** canoe slid on the smooth planks.
Base sentence	**Mary** said "Dagada" today.
Harvard sentence	Glue the **sheet** to the dark blue background.
Harvard sentence	It's **easy** to tell the depth of a well.
Base sentence	**Peter** said "doodads" today.

Table 1. Sample of the base corpus.

<table>
<tr><th colspan="2">General Scenario</th></tr>
<tr><td colspan="2">You are at a party for your friends Bobby and Francis, accompanied by your friend Julie. The subsequent scenarios are not necessarily interrelated.</td></tr>
<tr><td>Scenario</td><td>Sentence</td></tr>
<tr><td>Julie asks you, "Who gave Bobby doodads?" You reply:</td><td>Dagada gave Bobby doodads.</td></tr>
<tr><td>Julie asks you, "Did Dagada give doodads to Francis or Bobby?" You reply:</td><td>Dagada gave Bobby doodads.</td></tr>
<tr><td>Julie asks you, "What did Dagada give Bobby?" You reply</td><td>Dagada gave Bobby doodads.</td></tr>
<tr><td>You are not sure if Dagada gave Bobby doodads. You ask Julie:</td><td>Dagada gave Bobby doodads?</td></tr>
<tr><td>You are not sure whether Dagada gave Bobby or Francis doodads. You ask Julie:</td><td>Dagada gave Bobby doodads?</td></tr>
<tr><td>You are not sure whether or not Dagada gave Bobby doodads. You ask Julie:</td><td>Dagada gave Bobby doodads?</td></tr>
</table>

Table 2. Scenarios for the test sentences.

4.3 Selection of cycles for measurement

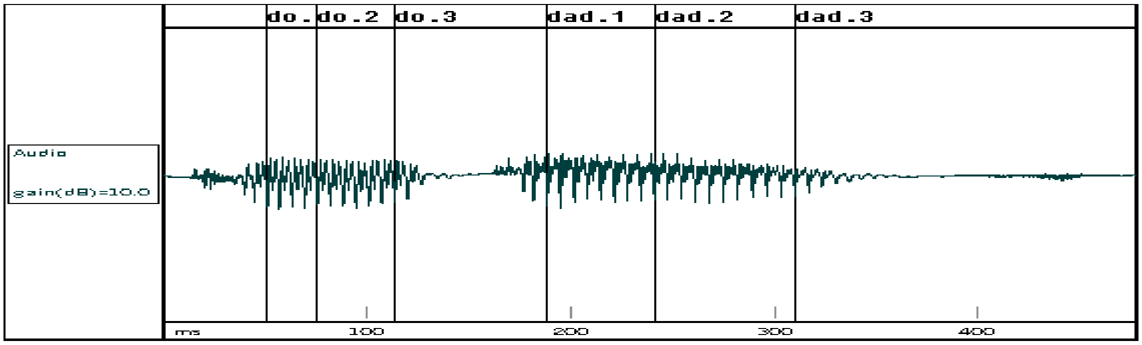
In this analysis, individual glottal pulses were selected for inverse-filtering and LF-fitting using signal analysis developed at UCLA's Bureau of Glottal Affairs.[5] A syllable-medial glottal cycle was selected from the vowel of each syllable of each word in the corpus. This cycle was selected from a steady portion of the vowel, i.e. the point at which one would ideally take spectral measurements. This was done to avoid as much as possible the transition between consonants and vowels. Syllable-initial and syllable-final cycles were also selected from *dads* in *doodads* and from each syllable on which speakers were asked to place narrow focus. Measurements were then averaged across samples from a single syllable (a pilot study had shown that measurements of a syllable-medial samples were not significantly different from syllable-edge samples). These additional samples will allow for better assessment of effects of prominence and sentence-final intonation contours on voice quality.

To avoid selecting cycles that are part of bursts or voiced consonants, the syllable-edge cycles were selected in the following manner (see Figure 25 for an example):

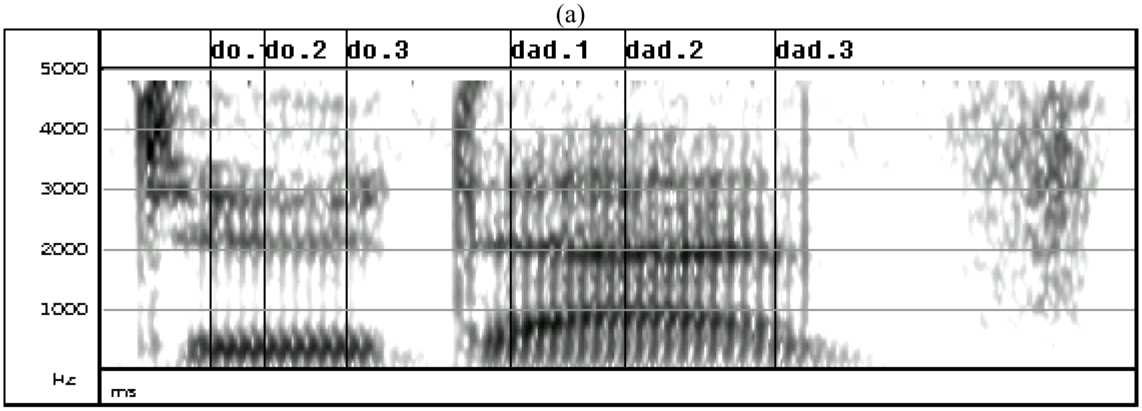
- At the beginning of the vowel the cycle should look like one cycle before and two cycles after.
- At the end of the vowel the cycle should look like two cycles before and one cycle after.

Cycles that could not be fit with the selected mathematical model (the Liljencrants/Fant (LF) model) were set aside and labeled as "not LF-fitable".

[5] Software for inverse-filtering and LF-fitting is available from the Bureau of Glottal Affairs on the following website: http://www.surgery.medsch.ucla.edu/glottalaffairs/software_of_the_boga.htm or by contacting Dr. Jody Kreiman at jkreiman@ucla.edu..



(a)



(b)

Figure 25. Acoustic waveform (a) and spectrogram (b) showing location of syllable-initial, -medial and –final cycles selected from focused and sentence-final syllables.

Since calculation of the vocal tract filter through LPC analysis requires a time window longer than a single cycle, cycles were concatenated with themselves 10 times at a zero crossing. This eased the estimation of the formants and bandwidths for the cycle selected for inverse filtering. A particular vocal tract filter was chosen in the following manner:

(1) The filter should be a reasonable estimation of the formants measured from:
–a spectrogram of the entire sentence.
–LPC and FFT analysis of the single, concatenated cycle.
(2) The output glottal flow and glottal flow derivative should have as few bumps and ripples as possible
(3) The spectrum of the glottal flow should be smoothly decreasing.

The glottal flow derivative, the output from the inverse filtering procedure, was then automatically fit with the mathematical model. Note that using measurements off of a single cycle to describe a series of cycles assumes that each cycle is more or less the same as the ones that precede and follow it. Syllables that did not contain a single characteristic glottal cycle were also set aside as not LF-fitable. Approximately 10% of the more than 1000 cycles selected for analysis could not be fit with the LF model.

4.4 Calculation of the LF model

Selected cycles were then fit with the LF model of differentiated glottal flow. As noted above, the LF model of the glottal flow derivative used in this study has been modified in order to better fit non-modal phonation, particularly in the modeling of the return phase (see Figure 26). The LF model in this study is composed of a sinusoid, an exponential and a linear term. The linear term is added to the exponential segment to force the return phase back to zero.

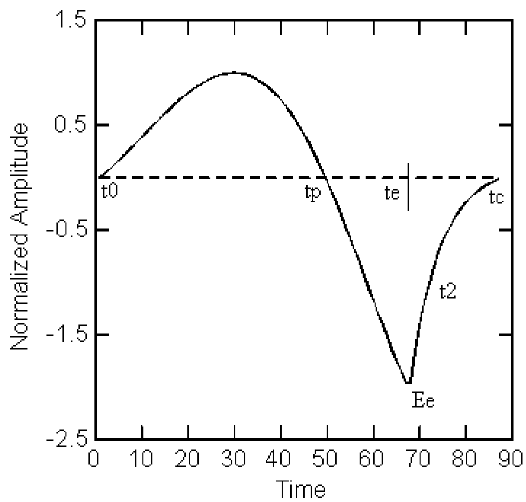


Figure 26. Revised LF model of differentiated glottal flow. Note that t_2 has replaced t_a.

The points used in calculation of the LF model are the following major features of the glottal flow derivative (U′) (see Figure 14):

t_0 = the start of the pulse
t_c = the length of the entire pulse
t_p = the length of the time that $U' > 0$
t_e = the time of the maximum negative value of U′
E_e = the value of the maximum negative U′
t_2 = the time increment to 50% decay in the return phase.

The sinusoid is calculated using the formula $U'(t) = E_0 \cdot e^{\alpha t} \cdot \sin(\omega_g t)$ with the following three parameters: $\mathbf{E_0}$ is a scale factor; $\boldsymbol{\alpha} = -B\pi$, where B is the "negative bandwidth" of the sinusoid, so the larger α, the faster the increase in amplitude; $\boldsymbol{\omega_g} = 2\pi F_g$, where $F_g = 1/(2t_p)$ and $\mathbf{t_p}$ is the time from glottal opening to maximum airflow. E_0, α and ω_g are calculated from t_p, t_e and E_e via non-linear solution. The exponential is calculated with the revised parameter $\mathbf{t_2}$. It has the formula $U'(t) = -E_e e^{-\varepsilon(t-te)} + m(t-t_e)$ with $\varepsilon = \ln(2)/(t_2 - t_e)$ and $m = (E_e e^{-\varepsilon(tc-te)})/(t_c - t_e)$. The "sixth" LF parameter is $\mathbf{F_0}$, the fundamental frequency. The model also assumes that the end of one pulse is the beginning of the next pulse, in other words, $t_{c(\text{pulse } 1)} = t_{0(\text{pulse } 2)}$. The values for all the parameters are calculated by a least squares minimization process on the time domain LF curve. Differences in LF parameters will also reflect differences in the shape of the glottal flow derivative. See Kreiman et al. (2000) for a detailed review of the calculation of the revised version of the LF model.

4.5 Not LF-fitable cycles

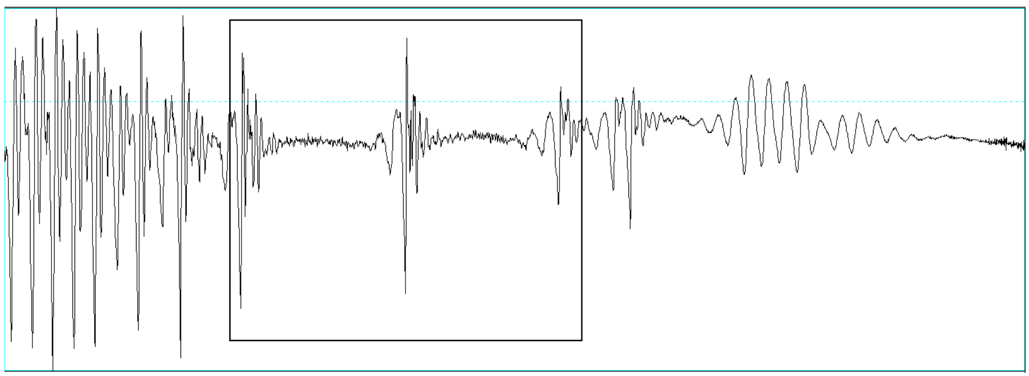
As mentioned above, cycles selected for measurement were placed into two categories:

1) LF-fitable, which were then further assessed with the LF model of the glottal flow derivative.
2) Not LF-fitable.

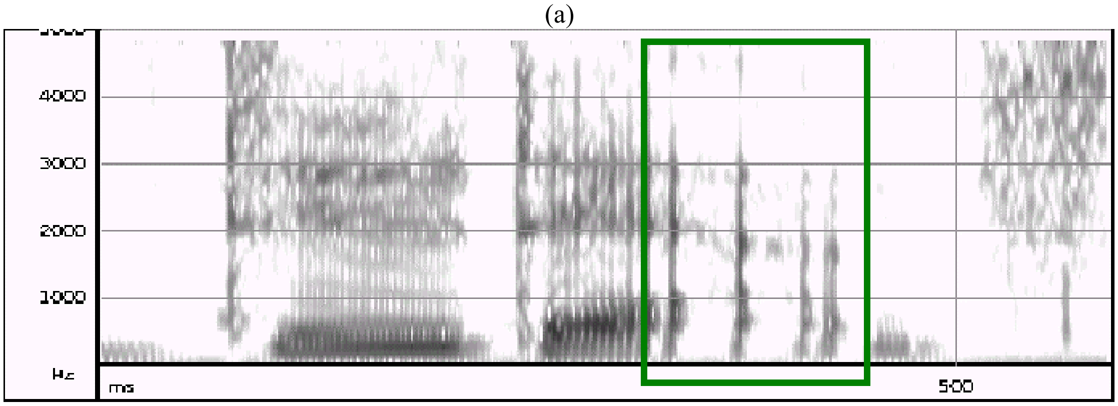
The LF model of glottal flow assumes that the source is periodic. There are many types of voice qualities, though, that do not have a periodic source (i.e. the vocal folds are vibrating irregularly or not at all). Such qualities occurred in approximately 10% of the total corpus and included some of the following voice qualities:

1) Vocal fry

Vocal fry is characterized by very low frequency (usually under 100 Hz), highly damped glottal pulses. The glottal pulses are often irregularly spaced. Vocal fry usually sounds like a series of separate taps, as opposed to a continuous, smooth vowel sound. In the literature it has also been referred to as creak and laryngealization. See Laver (1980) for a review.



(a)

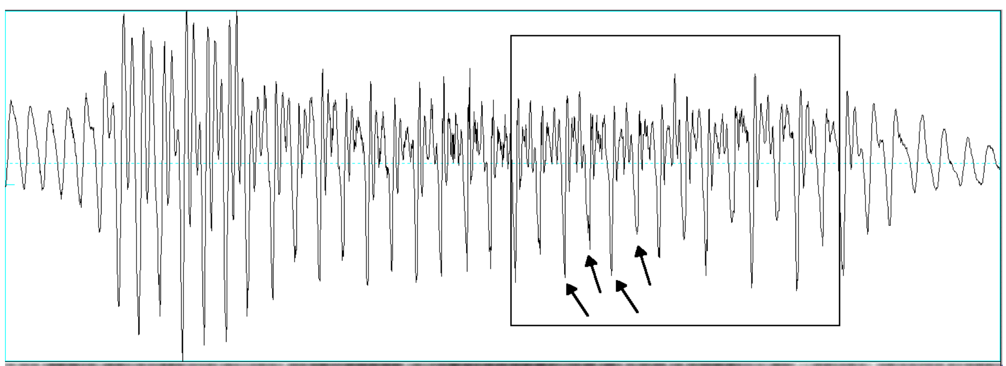


(b)

Figure 27. Acoustic waveform (a) and spectrogram (b) of a token with vocal fry. Note the highly damped, widely spaced glottal pulses in the waveform in (a). Note the irregularly spaced vertical striations in the spectrogram in (b).

2) Period doublings/Pitch halvings

Period doubled phonation has a repeating pattern that extends over more than one cycle of vocal fold vibration (see Figure 28). The periods can alternate in duration, amplitude or both. In the literature this kind of phonation has been referred to as bicyclicity, creak, diplophonia, etc. Period doubling usually sounds creaky like vocal fry, but not as low pitched. (See Gerratt & Kreiman (2001) for review). In English, period doubling has been associated with shifts between lower and high register in young children (Keating 1980). The doubled pulse cannot be fit with a single LF pulse model. Although each pulse in period doubled phonation may be inverse-filtered and fit with the LF model separately, this would not capture the true character of the phonation type, since the pulses are acting as a pair.



(a)

Figure 28. Acoustic waveform of a token with period doubling. Arrows show the alternating large-small cycles.

3) Aperiodic phonation

Aperiodic phonation is characterized by a series of glottal pulses that change rapidly in frequency and/or amplitude (see Figure 29). Although each individual glottal pulse may be inverse-filtered and fit with the LF model, an individual pulse would not be characteristic of the sample from which it is taken.

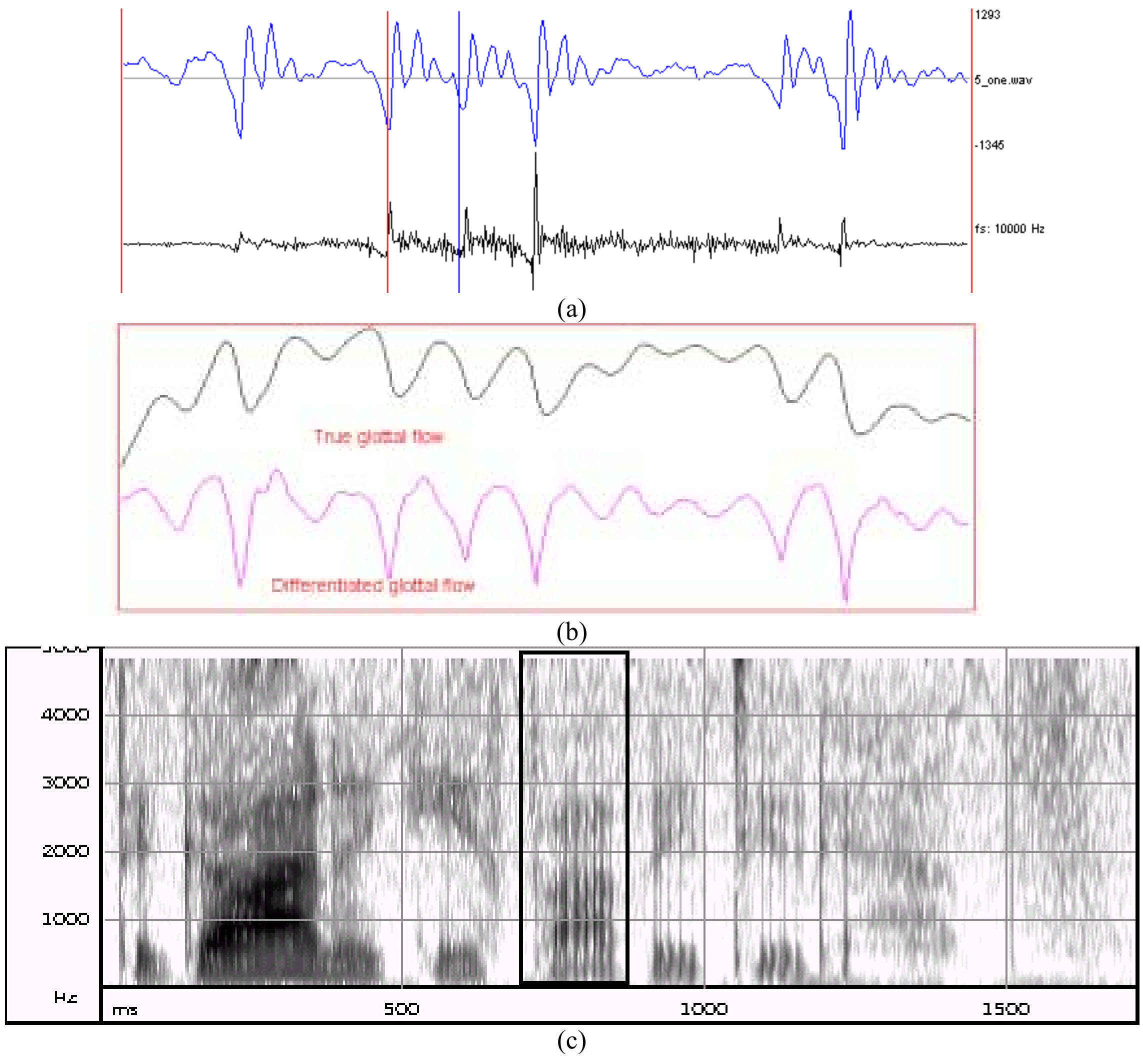
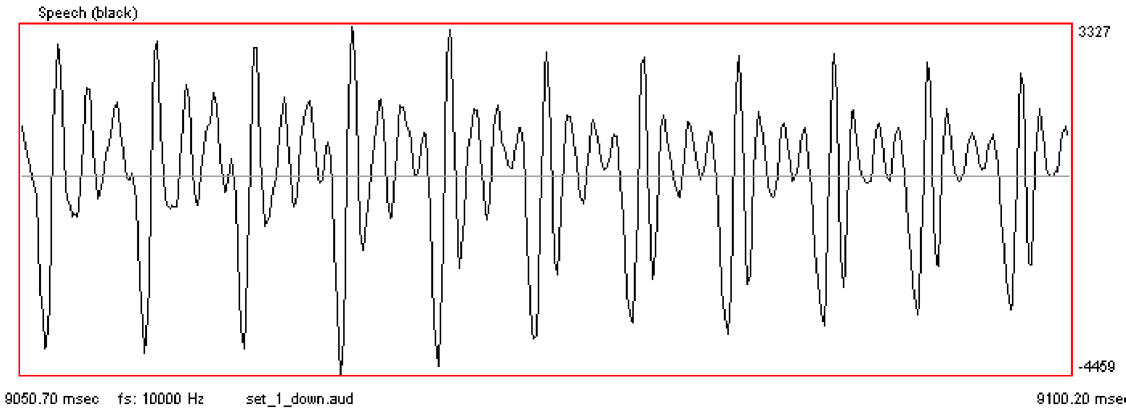
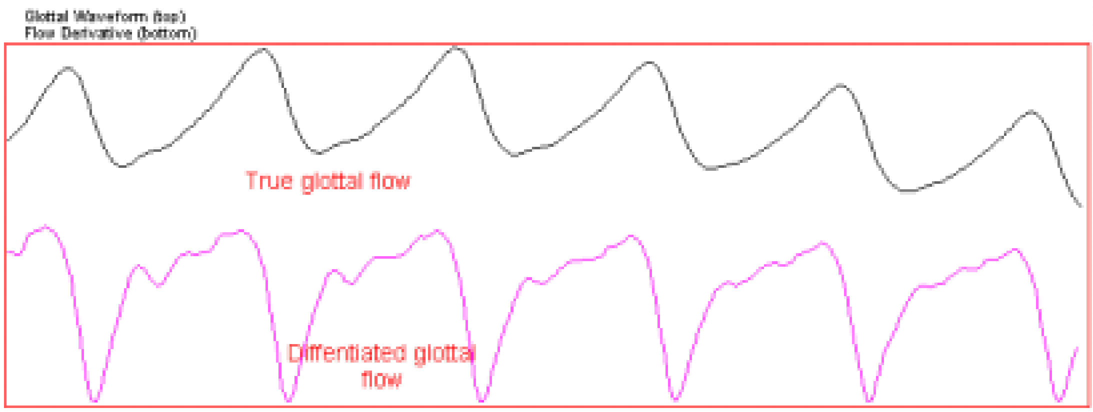


Figure 29. Acoustic waveform (a), glottal flow (b) and spectrogram (c) of a token with aperiodic phonation. Note the irregular spacing of the cycles in (a), and the matching irregularly spaced and shaped true and differentiated glottal flow in (b). Contrast with the waveform and glottal flow in Figure 30. Note the irregularly spaced striations in the boxed portion of the spectrogram in (c).



(a)

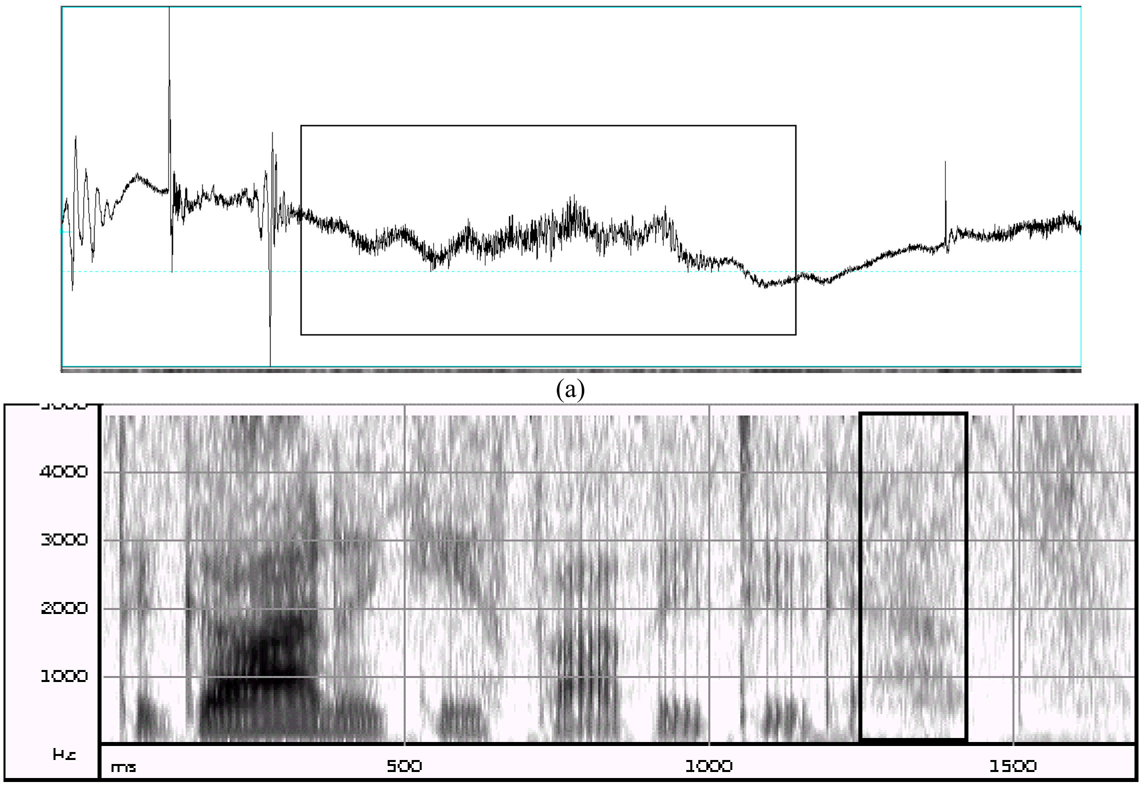


(b)

Figure 30. Acoustic waveform (a) and glottal flow (b) from a modal token.

4) Noise

Noise is characterized by complete lack of vocal fold vibration or supraglottal turbulence. The turbulent airstream, though, may still excite the vocal tract, as can be seen by the faintly visible formants in the boxed portion of the spectrogram in Figure 31.



(a)

(b)

Figure 31. Acoustic waveform and spectrogram of noise.

4.6 *Prosodic labeling*

The corpus was prosodically labeled so that comparisons can be made across different prominent positions, prominent and non-prominent words and different pitch accent and boundary tones. The labeling system was closely based on the ToBI (Tones and Break Indices) transcription standard (Beckman & Ayers 1997). It will be called Simple Tones to distinguish it from the ToBI system. The goal of the ToBI system is to be able to describe all of the categorically distinct intonation patterns in a language. It does not describe the continuous aspects of intonation, such as local changes in tempo or pitch, nor does it describe the categorical aspects of intonation that are predictable, such as lexical word stress. Examples of all sentence types used in this analysis, labeled with both the ToBI and Simple Tones systems, are provided in Figure 32, Figure 33,Figure 34 and Figure 35.

In the ToBI system, tones that are used to mark metrically strong syllables are called pitch accents (marked by a "*") and tones that are used to mark the edges of larger prosodic groupings are called boundary tones. In English, there are two levels of phrase tones. The intermediate phrase contains at least one accented syllable and the phrase accent (marked by "-") which marks the syllables between the last pitch accented syllable and the end of the phrase. The intonation phrase is marked with a boundary tone (marked by a "%") and contains at least one intermediate phrase. In the ToBI system the last pitch accent in an intermediate phrase is called the nuclear pitch accent and is perceived as the most prominent accent of the phrase.

Both Simple Tones and ToBI assign accent and boundary tones according to phonological categories and not phonetic realization of the pitch track. For example, Figure 34 and Figure 35 contrast the sentences ***Dagada*** *gave Bobby doodads?* and *Dagada gave Bobby* ***doodads****?* spoken by the same speaker. In both tokens, the word *Dagada* is pronounced at ~200 Hz. Although, the word *Dagada* in ***Dagada*** *gave Bobby doodads?* (Figure 34) is perceived as lower than the following words, it is labeled as a High tone in both systems. This is because this sentence can be compared with the sentence *Dagada gave Bobby* ***doodads****?* (Figure 35) where *Dagada* is also produced at ~200 Hz and also is labeled with a High tone. The *Dagada*s can also be contrasted with the *doodads* in shown Figure 35, which is produced on a much lower pitch, ~170 Hz, and is labeled with a Low tone. In other words, knowing the speaker's pitch range allows labelers to label both *Dagada*s in Figure 34 and Figure 35 with High tones, and *doodads* in Figure 35 with a Low tone.

The primary difference between Simple Tones and ToBI is that Simple Tones collapses all of the ToBI tones into four categories – Low (L), Fall (F), High (H) and Rise (R). In the analysis of voice quality done for this study, the four categories are then further collapsed into two categories, Low (low and fall) and High (high and rise) on the basis of the tonal goal of the pattern. At the end of an intonation phrase, the Simple Tones system also collapses together the intermediate phrase accent and the intonation phrase boundary tone into a single boundary tone. These ToBI categories are collapsed to allow for comparison of voice qualities across broad, intuitive categories. For example, a ToBI H- phrase accent plus an L% boundary tone would be used to describe

the high level contour at the end of the sentence. The Simple Tones system labels this as a high boundary tone.

Moreover, Simple Tones, unlike ToBI, does not require labelers to label pitch accents, phrase accents or boundary tones to explain the visible pitch contour. For example, the Simple Tones system transcribes “prominent” accents that are defined as the “most prominent” pitch accent(s) in a phrase. A prominent accent does not have to be the last pitch accent within an intermediate phrase. One or more regular pitch accents may follow. Also, labelers may label one or more pitch accents as being equally “most prominent” after the prominent accent. In Simple Tones, “prominent” pitch accents are labeled by a double (**) and non-prominent pitch accents by a single star (*). The Simple Tones system does not require that there be an intermediate phrase boundary between prominent accents, unless the labeler perceives a boundary due to a break, phrase-final lengthening or significant change in pitch. For example, ***Dagada*** *gave Bobby doodads.* in Figure 32 is identified as having two prominent pitch accents under both the ToBI and Simple Tones systems. Under the ToBI system, the labeler must insert a low intermediate phrase boundary between the two nuclear pitch accents. In the Simple Tones system, both prominent accents may exist in the same perceptual intermediate phrase.

It is worth reiterating that the Simple Tones system was designed purposely for this small, highly constrained corpus. It does not adequately handle all intonation features of English.

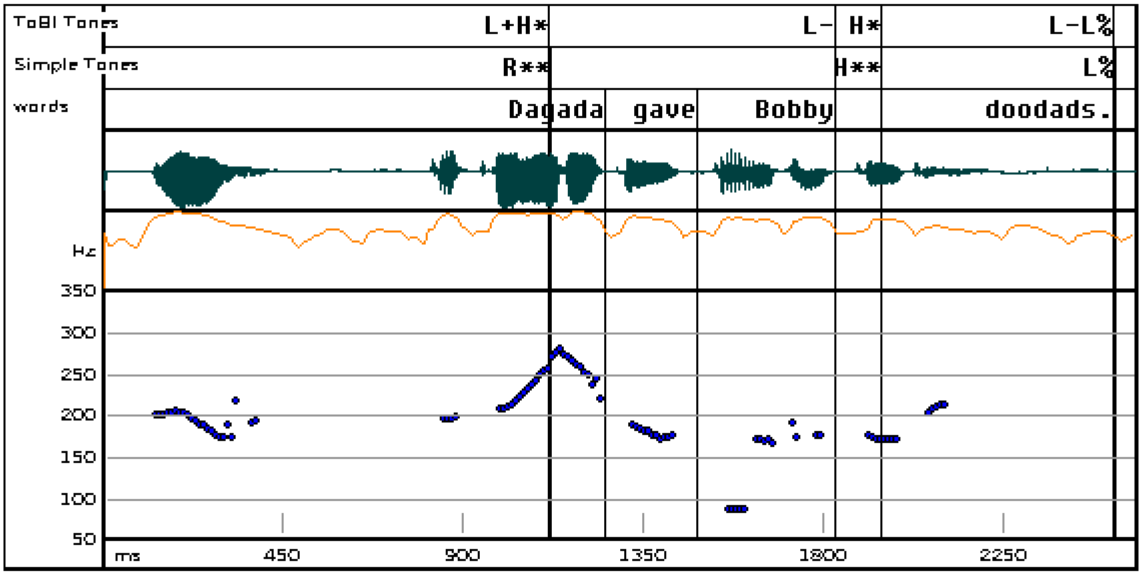


Figure 32. Pitch contour, ToBI and Simple Tones labels for ***Dagada*** *gave Bobby doodads.* as spoken by subject B. ToBI tones are in the top row, Simple tones are in the middle row. This speaker has a slight rising contour for her low boundary tones in some declarative sentences.

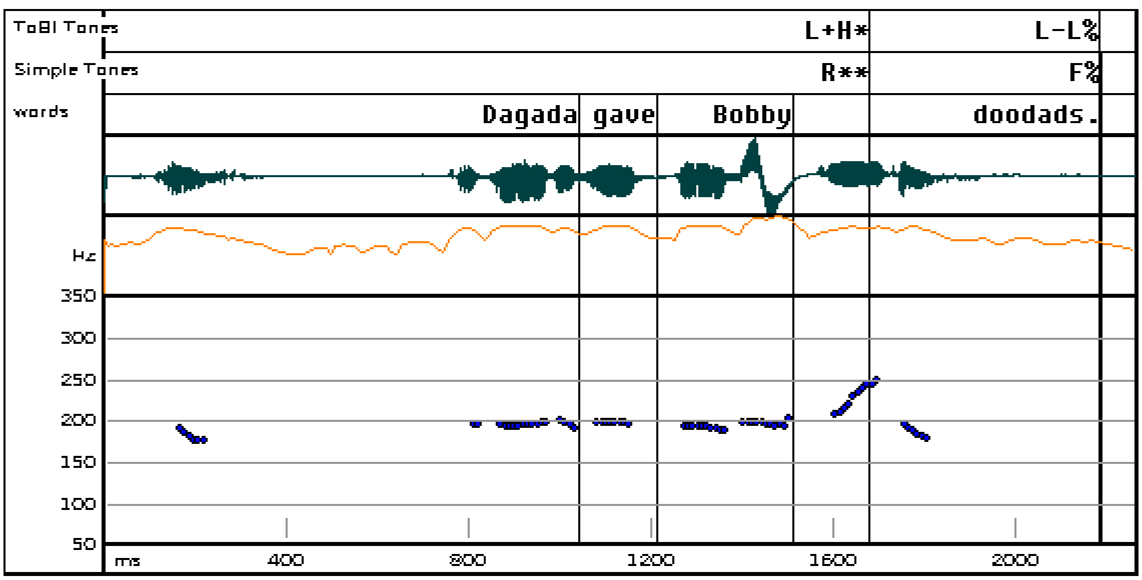


Figure 33. Pitch contour, ToBI and Simple Tones labels for *Dagada gave Bobby* ***doodads***. as spoken by subject B. ToBI tones are in the top row, Simple tones are in the middle row.

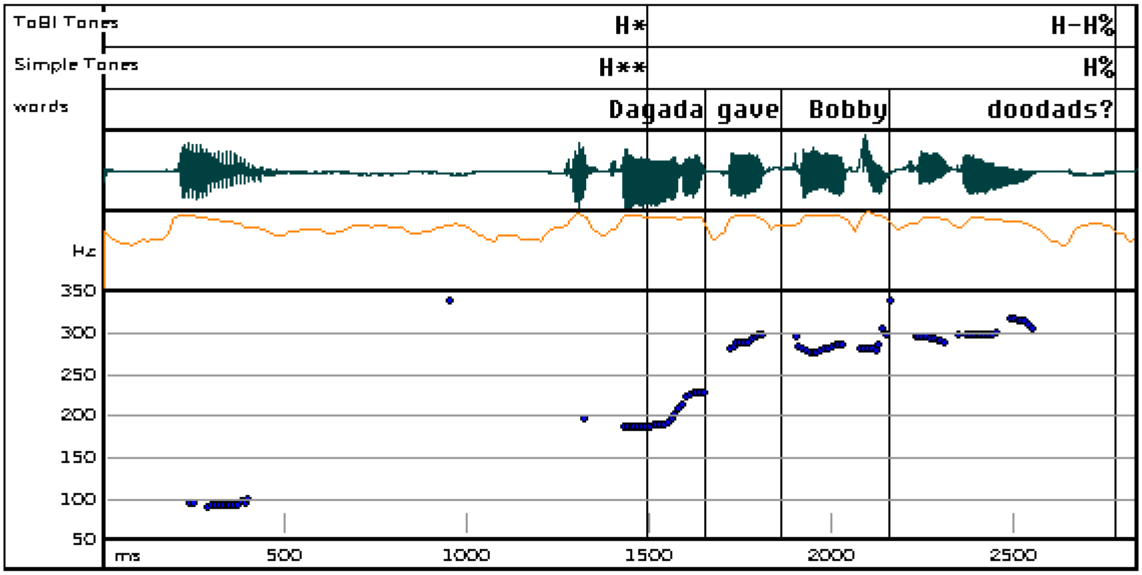


Figure 34. Pitch contour, ToBI and Simple Tones labels for ***Dagada*** *gave Bobby doodads?* as spoken by subject B. ToBI tones are in the top row, Simple tones are in the middle row.

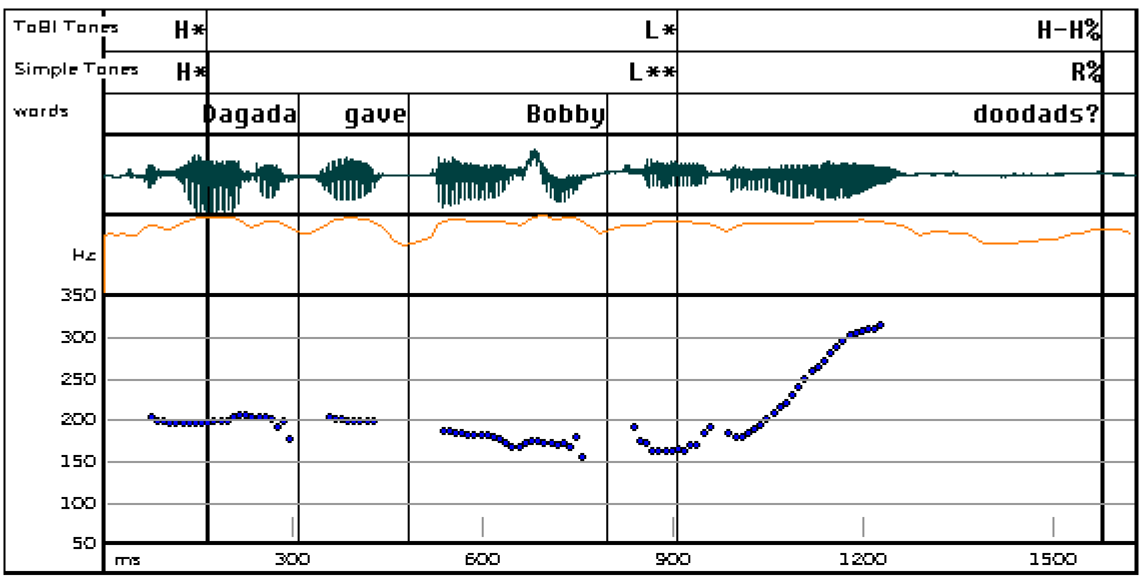


Figure 35. Pitch contour, ToBI and Simple Tones labels for *Dagada gave Bobby* ***doodads****?* as spoken by subject B. ToBI tones are in the top row, Simple tones are in the middle row.

4.7 Description of the test corpus used for analysis

The final test corpus consisted of cycles selected from the following four sentences:

- **Dagada** gave Bobby doodads.
- Dagada gave Bobby **doodads**.
- **Dagada** gave Bobby doodads?
- Dagada gave Bobby **doodads**?

Although the speakers repeated each set of sentences 10 times, only at the 8 medial repetitions were selected for further analysis. There were then 96 sentences total in the test corpus (3 speakers by 4 sentences types by 8 repetitions). Two expert ToBI labelers were trained by the author in the Simple Tones system. They then labeled the test corpus independently and then compared their labels to create a unified analysis. In their analysis they did not identify any intermediate phrases. For three sentences their prominent accent label did not fall on the word the speaker had been asked to narrowly focus.

The distribution of the different prominence levels (e.g., not prominent, pitch accented and prominent) is described in Table 3. As can be seen there is not an even distribution of prominence levels throughout the corpus. In particular, speakers had a tendency to pitch accent the first word, *Dagada*, when they made the final word, *doodads*, prominent. However, the number of prominent accented tokens for phrase-initial *Dagada* (59) and phrase-final *doodads* (47) is not significantly different ($p \leq 0.285$ for a two-tailed binomial test).

For the analysis, High (H) and Rise (R) pitch accent and boundary tones were combined into a single High tone category for pitch accents and boundary tones and

Falling (F) and Low (L) boundary tones (there were no F pitch accents) were combined into a single Low tone category (boundary and pitch accent tones were still kept separate). All interrogatives ended with a High boundary tone and all declaratives ended with a Low boundary tone. Table 4 shows that there was not an even distribution of High and L tones in the corpus; there were significantly more High tones ($p \leq 0.003$ for a two-tailed binomial test). This is due to speakers using High prominent pitch accents in declaratives and even in interrogatives; Low prominent pitch accents, however, only appeared in interrogatives.

		Word				
		Dagada	gave	Bobby	doodads	**Total**
Prominence Level of Word	not prominent	2	96	88	45	231
	pitch accented	35		8	4	47
	prominent accent	59			47	106
	Total	96	96	96	96	384

Table 3. Distribution of prominence levels (not accented, pitch accented and prominent) in the test corpus.

		Tune		
		declaratives	interrogatives	**Total**
Prominent Pitch Accent Tone	H	55	14	69
	L		37	37
Total		55	51	106

Table 4. Distribution of the type of prominent pitch accent tones in the test corpus.

There was also an uneven distribution among the speakers within the test corpus of not-LF-fitable waveforms. Speaker B produced the greatest proportion of non-modal phonation in the test corpus at 16% (chi-square = 27.455, $p \leq 0.001$), as seen in Figure 36.

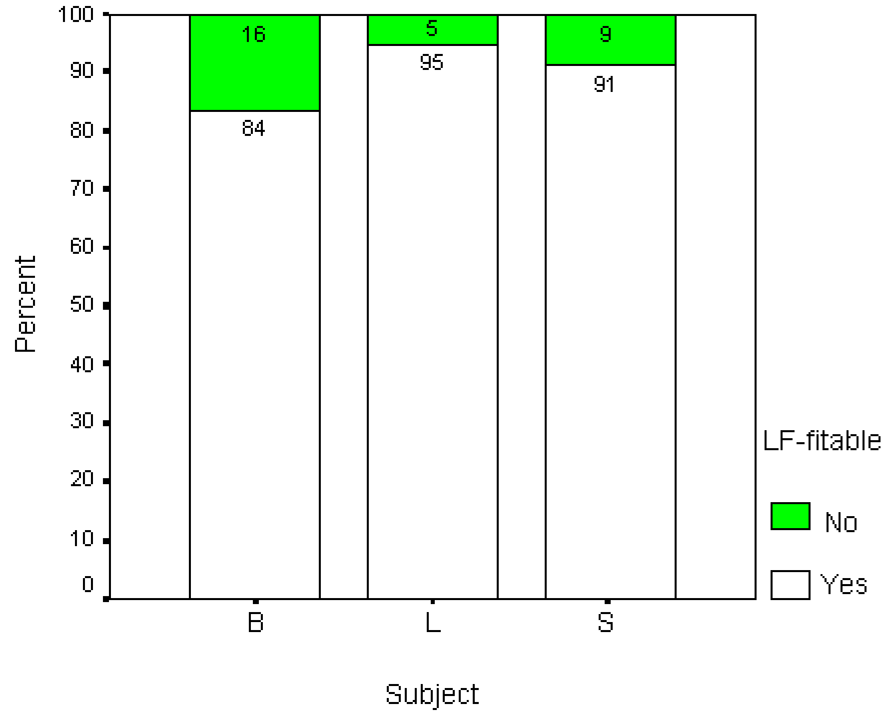


Figure 36. Proportion of not-LF-fitable waveforms in the test corpus for each subject.

4.8 Statistical methods

Evaluation of the effects of prosody on voice quality is based on the General Linear Model. Specific tests included Multivariate Analyses of Variance (MANOVAs) and Analyses of Variance (ANOVAs). All statistic analysis was conducted using the SPSS statistical package (Version 10, SPSS Inc., 1999). Although the data contain multiple repetitions of the test sentences by the subjects, a repeated measures design was not used because there are non-predictable categorical independent factors resulting from the prosodic labeling of the corpus. In other words, the locations of prominent tones and the type of prominent and boundary tones was not completely predictable from the structure of the corpus. As a result both repetitions and between-subject interactions are a source of error in the statistical model.

Due to the exploratory nature of this study, and to decrease the probability of a Type II error, a fairly high experimental alpha is chosen. This decreases the probability of falsely accepting the null hypothesis that prosody has no effect on voice quality. Thus, each family of statistical tests had an experimental alpha level of 20%. The alpha is then adjusted to account for the number of univariate ANOVAs conducted or the number of dependent variables compared in a correlation. For the correlations, five dependent factors are compared (EE, RK, OQ, Linearity and F0), so each correlation is judged significant at $p \leq 0.04$. The ANOVAs testing the univariate effects for the four dependent factors EE, RK, OQ and Linearity and the between-subjects effects are judged significant at $p \leq 0.05$.

The independent factors considered in this study are Prominence (Prominent, Non-Prominent), Phrase Edge (Initial, Final) and Tone (High, Low). The data used for each factor in the statistical tests are summarized in Table 5, Table 6 and Table 7. The samples associated with each factor were selected so that no test levels were weighted more strongly than others.

	Prominence			
Tests	Word Prominence		Stressed Syllable Prominence	
Levels	Prominent	Not	Prominent	Not
Sample(s)	syllable-medial samples for stressed and unstressed syllables (2-3 samples depending on syllables/word)		syllable-medial samples	

Table 5. Samples used in analysis of prominence.

	Phrase Edges			
Tests	Prominent Words		Large Sections	
Levels	Initial	Final	Initial	Final
Sample(s)	syllable-initial, -medial & -final samples for stressed syllables; syllable medial samples for unstressed syllables		all samples in *Dagada gave* (4 – 6 samples, depending on focus)	all samples in *Bobby doodads* (4 – 6 samples, depending on focus)

Table 6. Samples used in analysis of phrase edges.

	Tone					
Tests	Prominent Syllables		Prominent Words		Final Syllable	
Levels	High	Low	High	Low	High	Low
Samples(s)	syllable-initial, -medial & -final samples		syllable-initial, -medial & -final samples for stressed syllables; syllable medial samples for unstressed syllables		syllable-initial, -medial & -final samples	

Table 7. Samples used in analysis of phonological tone.

The results are also presented graphically by means and 95% confidence interval bars clustered by individual subjects. These assist in assessing the effects of prosody on the

voice quality for individual subjects and the between-subjects effects. As it turns out, the 95% confidence interval bars do not differ very much from bars representing standard error. Therefore, when the bars do not overlap, the difference between the means can be judged significant at the 95% confidence level and when the error bars *do* overlap the difference in the means is *not* significant at this level.

Chapter 5: Results

In review, this study uses four normalized measurements of the LF model of differentiated glottal flow to assess the effects of prominence, phrase edges and tone on voice quality in English. As a reminder, these parameters are:

- **EE** (spectral intensity) – the length of the negative peak of the differentiated glottal pulse. High values of EE are correlated with an increase in amplitude of all the harmonics in the glottal source spectrum and with a tenser voice quality.
- **Linearity** (spectral linearity) – how closely a line fit to the spectrum of the glottal pulse matches a straight line. A high value of Linearity is associated with a glottal spectrum that has a low spectral tilt and/or few spectral notches. High values of Linearity are correlated with a tenser voice quality.
- **RK** (glottal symmetry/skew) – the ratio of the closing phase to the opening phase of the differentiated glottal pulse. Low values of RK are correlated acoustically with fewer spectral notches and higher amplitude high frequency harmonics and with a skewed glottal pulse. Low values of RK are associated with a tenser voice quality.
- **OQ** (open quotient) – the ratio of the opening phase to the period of the differentiated glottal pulse. Low values of OQ are correlated acoustically with higher amplitude high frequency harmonics. Low values of OQ are associated with a tenser voice quality.

In short it is expected that a tense voice quality will have large values for EE and Linearity, and small values for RK and OQ. Lax voices, on the other hand, will have small values for EE and Linearity, and large values for RK and OQ.

As discussed in Chapter 3, all measurements in this study are normalized to a baseline recording for each speaker. The average of the baseline values is shown as the zero line in figures below. The base sentences were:

- **Mary** said “Dagada” today.
- **Nancy** said “Bobby” today.
- **Dana** said “gave” today.
- **Peter** said “doodads” today.

No patterns could be generalized, however, by making a comparison to the "zero line". This is probably due to the three speakers producing somewhat different intonation patterns for their base sentences. It is important to note though, that the three speakers did accomplish the two primary goals of internal consistency and not producing an extreme pitch excursion on the base word. Speakers B and L both had an intermediate phrase break after the boldfaced, accented name, e.g. the *Mary* in "**Mary** said 'Dagada' today." Speaker B had a ToBI (L+)H* accent for the base word (e.g. 'Dagada') and Speaker L had a ToBI H* accent for the base word. Speaker S did not produce an intermediate phrase break in the sentence, so her base words were deaccented. They were, however, fully articulated. In short, the baselines did accomplish the task of factoring out segmental and personal voice quality differences in the test corpus, but it did not create a base *line* against which comparisons could be made.

Results will now be presented in four sections:

- Correlations among normalized LF measurements and F0
- Effects of prominence on voice quality
- Effects of prosodic domain edges on voice quality
- Effects of phonological tone on voice quality

5.1 Correlations among F0 and the LF measurements

Changes in F0 may be correlated with changes in voice quality and/or vocal fold vibrations, since a high F0 is correlated with increased vocal fold tension and length and a low F0 is correlated with decreased vocal fold tension and length, as reviewed in section 2.3.1 (Laver 1980; Ohala 1973). To find out if measured F0 is correlated with the voice quality measures used for this corpus, Pearson product moment correlation coefficients were calculated for all five normalized measurements of the LF pulse: F0 (phonetic pitch), EE (spectral intensity), RK (glottal symmetry/speed of glottal closure), OQ (open quotient/glottal closing duration), Linearity (spectral linearity). All LF-fit cycles in the corpus were included in the analysis. Correlations were judged to be strong if r was greater than or equal to 0.65.

For all the speakers as a group, it was found that F0 did not correlate strongly with any of the LF measurements (see Table 8). In other words, as F0 increased or decreased, no other LF measurement correspondingly increased or decreased. For the speakers as individuals, it was found for both speakers B and S that F0 did not correlate strongly with any of the LF measurements (see Table 9 and Table 10). For speaker L, on the other hand, it was found that two measurements correlated strongly with F0: Lin ($r = 0.683$, $p \leq 0.001$) and RK ($r = -0.663$, $p \leq 0.001$) (see Table 11). This indicates that as speaker L's pitch increased, his glottal pulses had greater values of glottal skew and his glottal spectra had stronger high frequency harmonics (lower values of spectral tilt). Closer inspection of the scatterplot of Glottal Skew (RK) against F0 for speaker L (see Figure 38) reveals, however, that high values of normalized F0 all have a normalized RK value of

approximately -0.05, and that high values of normalized RK all have a normalized F0 of approximately -0.2. In other words, there is not a linear relationship between normalized F0 and normalized RK for speaker L.

In summary, although F0 and voice quality may be correlated in some phonation type languages, such as Hindi, there is not a strong correlation between F0 and voice quality for the three speakers in this study. This result supports the results of previous studies on the relationship between voice quality and pitch in English (Epstein & Payri 2001; Holmberg et al. 1995). Of course, no LF measurements could be made for non-LF-fitable cycles. The relationship between non-modal phonation and phonological tone will be discussed in section 5.4.4.

5.1.1 Other correlations

Even though there was not a strong correlation between voice quality and F0, it is likely that other LF measurements will be correlated with each other. This is because changes in any one parameter of the LF pulse affect the shape, and thus the acoustic properties, of the pulse as a whole. It was found across all the subjects that spectral linearity (Linearity) and spectral intensity (EE) are positively correlated ($r = 0.747$, $p \leq 0.001$). Linearity is also somewhat less strongly negatively correlated with RK (speed of vocal fold closure) ($r = -0.641$, $p \leq 0.001$). See Figure 39 for the scatterplots. These correlations indicate that as the signal becomes more intense and as the glottal pulse becomes more skewed, there is a decrease in spectral tilt and spectral notches (*i.e.*, Linearity increases).

A review of the speakers as individuals shows that speaker B has a strong correlation of EE with Linearity (r = 0.651, $p \leq 0.001$) (see Figure 40) and speaker S has a strong correlation with Linearity for both EE (r = 0.673, $p \leq 0.001$) and RK (glottal skew) (r = -0.726, $p \leq 0.001$) (see Figure 41). For speaker L, EE and RK are strongly correlated with each other, (r = -0.690, $p \leq 0.001$) and Linearity and EE are strongly correlated with each other (r = 0.832, $p \leq 0.001$). As can be seen in Figure 42, though, there is not a linear relationship between EE and Linearity.

In summary, as expected, there are strong correlations among several of the LF measurements both for the speakers as a group and for the individual speakers. Because Linearity and EE are strongly correlated for the group as a whole and across all subjects, only EE will be used in the subsequent multivariate analyses.

All Speakers						
		Normalized F0	Normalized EE	Normalized RK	Normalized OQ	Normalized Linearity
Normalized F0	Correlation	1.000	.229	-.003	.285	.224
	p-value ≤	.	.001	.916	.001	.001
	N	1035	1035	1035	1035	1035
Normalized EE	Correlation		1.000	-.566	.019	.747
	p-value ≤		.	.001	.544	.001
	N		1035	1035	1035	1035
Normalized RK	Correlation			1.000	.068	-.641
	p-value ≤			.	.029	.001
	N			1035	1035	1035
Normalized OQ	Correlation				1.000	-.053
	p-value ≤				.	.086
	N				1035	1035
Normalized Linearity	Correlation					1.000
	p-value ≤					.
	N					1035

Table 8. Correlations of variables across all subjects. Correlations at or above 0.65 are highlighted.

Speaker B						
		Normalized F0	Normalized EE	Normalized RK	Normalized OQ	Normalized Linearity
Normalized F0	Correlation	1.000	-.429	.301	.050	-.192
	p-value ≤	.	.001	.001	.370	.001
	N	321	321	321	321	321
Normalized EE	Correlation		1.000	-.603	-.356	.651
	p-value ≤		.	.001	.001	.001
	N		321	321	321	321
Normalized RK	Correlation			1.000	.470	-.576
	p-value ≤			.	.001	.001
	N			321	321	321
Normalized OQ	Correlation				1.000	-.385
	p-value ≤				.	.001
	N				321	321
Normalized Linearity	Correlation					1.000
	p-value ≤					.
	N					321

Table 9. Correlations of variables for Subject B. Correlations at or above 0.65 are highlighted.

Speaker S						
		Normalized F0	Normalized EE	Normalized RK	Normalized OQ	Normalized Linearity
Normalized F0	Correlation	1.000	-.257	.320	.303	-.204
	p-value ≤	.	.001	.001	.001	.001
	N	350	350	350	350	350
Normalized EE	Correlation		1.000	-.576	-.232	.673
	p-value ≤		.	.001	.001	.001
	N		350	350	350	350
Normalized RK	Correlation			1.000	.179	-.726
	p-value ≤			.	.001	.001
	N			350	350	350
Normalized OQ	Correlation				1.000	-.148
	p-value ≤				.	.005
	N				350	350
Normalized Linearity	Correlation					1.000
	p-value ≤					.
	N					350

Table 10. Correlations of variables for Subject S. Correlations at or above 0.65 are highlighted.

Speaker L						
		Normalized F0	Normalized EE	Normalized RK	Normalized OQ	Normalized Linearity
Normalized F0	Correlation	1.000	.625	-.663	.304	.683
	p-value ≤	.	.001	.001	.001	.001
	N	364	364	364	364	364
Normalized EE	Correlation		1.000	-.690	.083	.832
	p-value ≤		.	.001	.115	.001
	N		364	364	364	364
Normalized RK	Correlation			1.000	-.313	-.660
	p-value ≤			.	.001	.001
	N			364	364	364
Normalized OQ	Correlation				1.000	.099
	p-value ≤				.	.058
	N				364	364
Normalized Linearity	Correlation					1.000
	p-value ≤					.
	N					364

Table 11. Correlations of variables for Subject L. Correlations at or above 0.65 are highlighted.

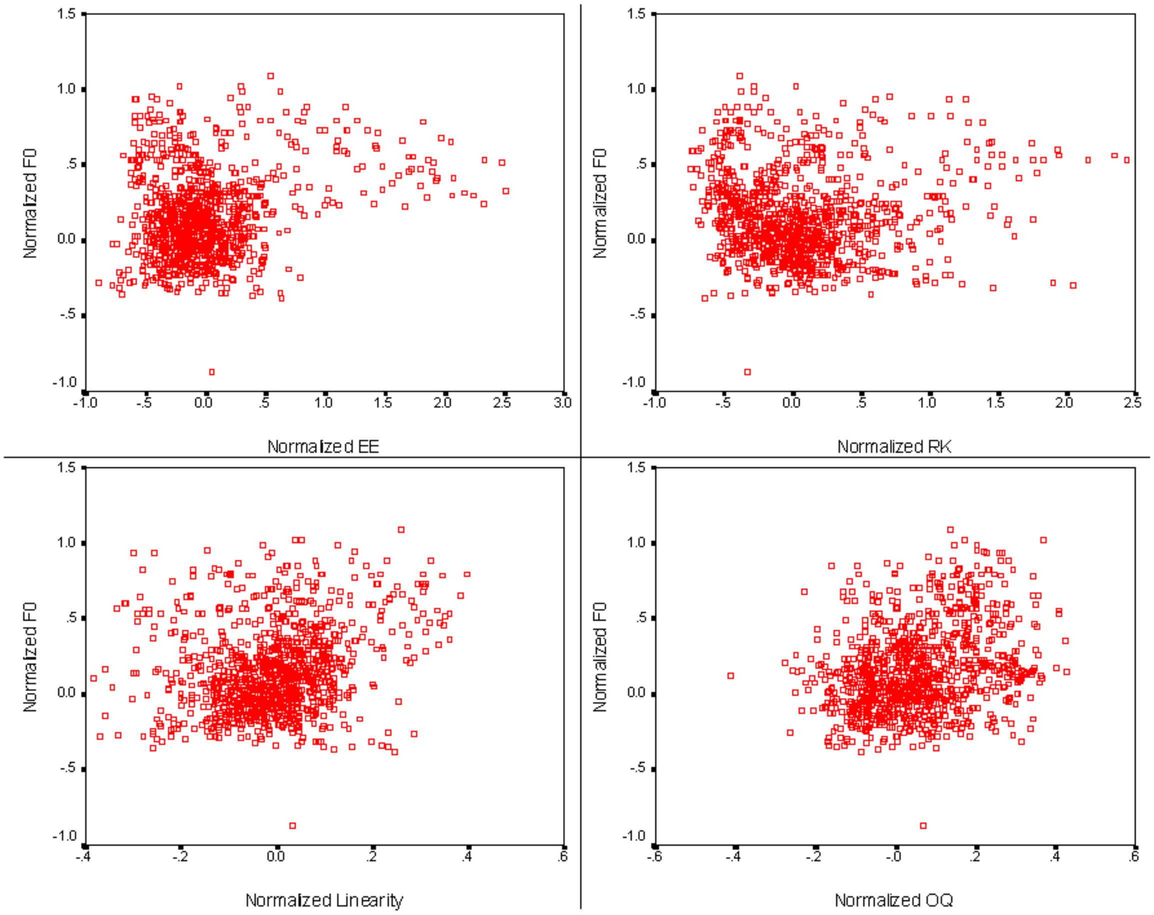


Figure 37. Scatterplots for correlations among F0 and the four LF measurements for all speakers combined: EE (spectral intensity), RK (glottal symmetry), OQ (open quotient) and Linearity (spectral linearity).

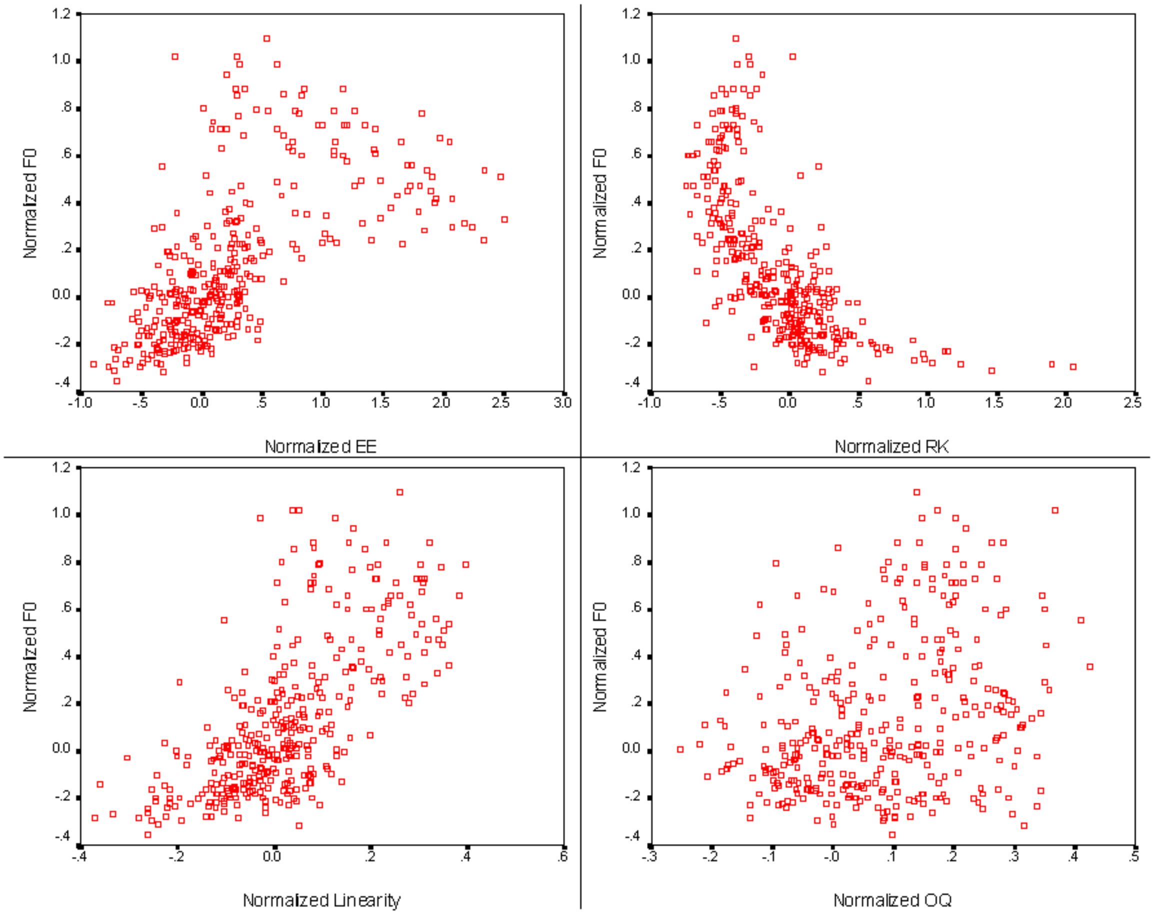


Figure 38. Scatterplots for correlations among F0 and the four LF measurements for speaker L: EE (spectral intensity), RK (glottal symmetry), OQ (open quotient) and Linearity (spectral linearity).

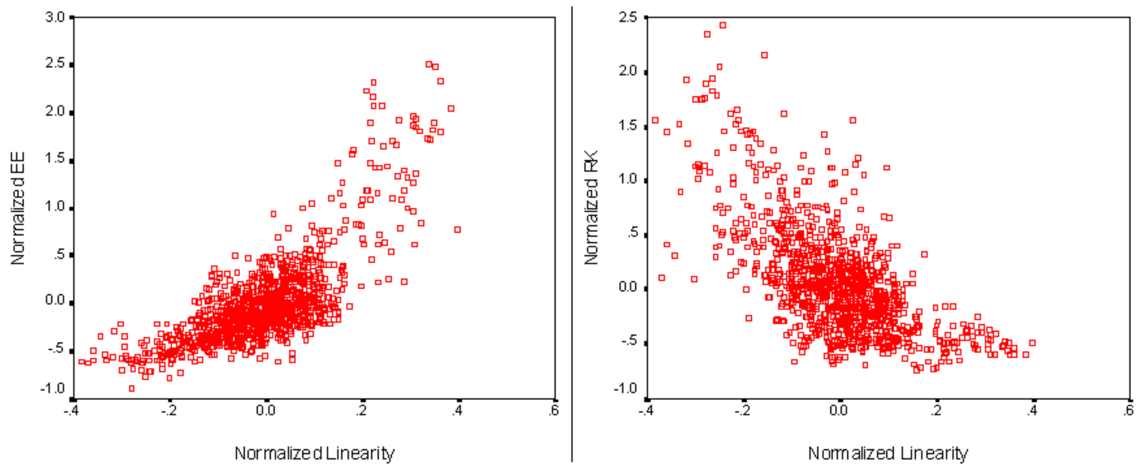


Figure 39. Scatterplots for significant correlations among the four LF measurements for all speakers grouped together: EE (spectral intensity), RK (glottal symmetry), and Linearity (spectral linearity).

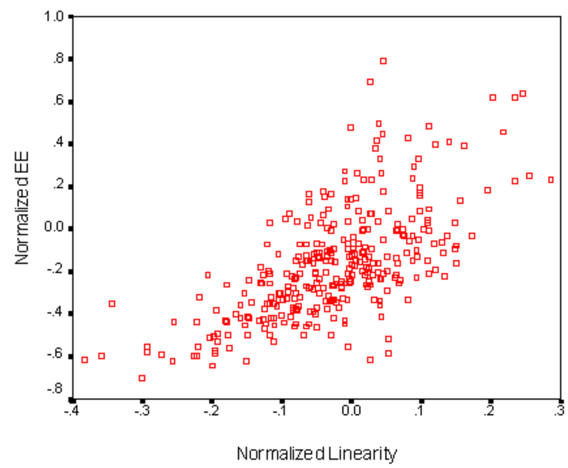


Figure 40. Scatterplot for significant correlations among the four LF measurements for speaker B: EE (spectral intensity) and Linearity (spectral linearity).

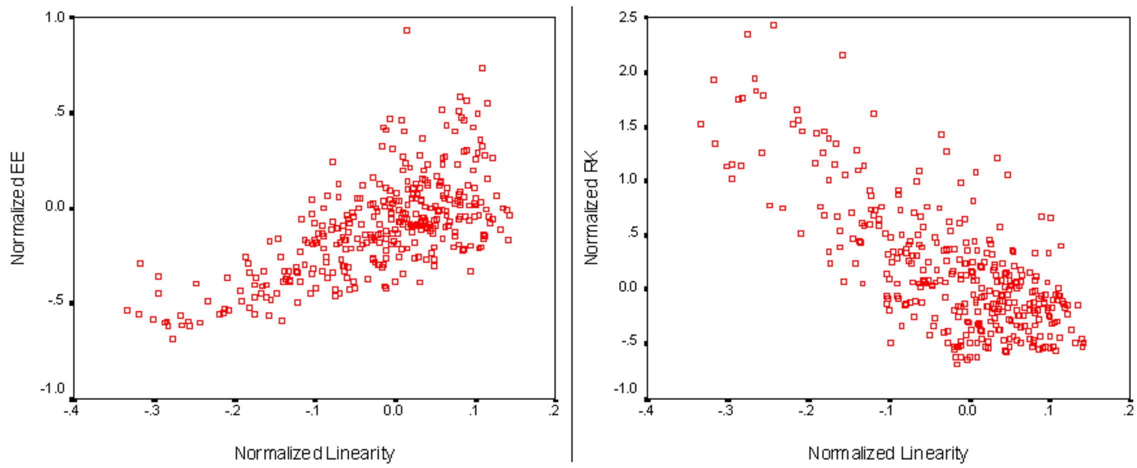


Figure 41. Scatterplots for significant correlations among the four LF measurements for speaker S: EE (spectral intensity), RK (glottal symmetry), and Linearity (spectral linearity).

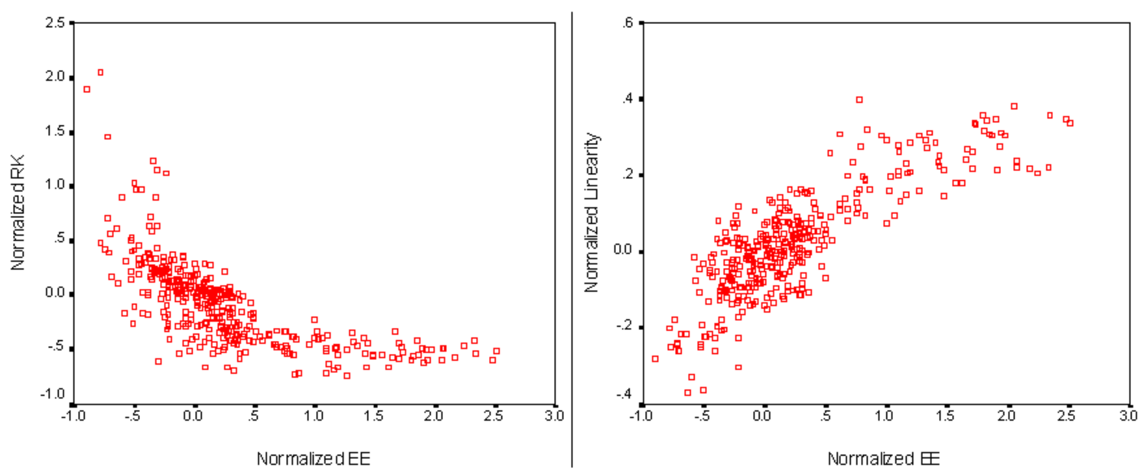


Figure 42. Scatterplots for significant correlations among the four LF measurements for speaker L: EE (spectral intensity), RK (glottal symmetry) , and Linearity (spectral linearity).

5.2 Prominence

Previous studies have shown that prominence affects the likelihood of vowel-initial glottalizations, as reviewed in section 2.3.2 (Dilley & Shattuck-Hufnagel 1995; Pierrehumbert 1995) and stressed or prominent words and syllables have a tenser voice quality (Gobl 1988; Campbell 1995). This section therefore examines the effects of prominence on the voice quality of words and stressed syllables for all sentence intonations combined. It also examines the effects of prominence on words in declarative and interrogative sentences separately to see if there is an effect of intonation tune. The following hypotheses are tested:

- **Prominent words have a tenser voice quality than non-prominent words for all sentence intonations combined.**
- **Prominent stressed syllables have a tenser voice quality than non-prominent stressed syllables for all sentence intonations combined.**
- **Prominent words have a tenser voice quality than non-prominent words in declarative sentences.**
- **Prominent words have a tenser voice quality than non-prominent words in interrogative sentences.**

5.2.1 Prominent vs. non-prominent words

A comparison of syllable-medial samples from prominent and non-prominent words found significant univariate effects for all four LF measurements for the three speakers combined. These results are summarized in Table 12. The MANOVA for the three speakers together did not find a significant main effect [F(1,572) = .177, $p \leq 0.674$]; there were, however, significant between-subject interactions for RK [F(2,568) = 4.762, $p \leq 0.009$].

Word		Normalized EE	Normalized RK	Normalized OQ	Normalized Linearity
non-prom	Mean	-.09860	.18210	.08457	-.03556
	N	323	323	323	323
	Std. Deviation	.41934	.61339	.13462	.12911
prominent	Mean	.14187	-.00084	.04531	.01707
	N	251	251	251	251
	Std. Deviation	.58567	.40263	.12830	.11769
	F(1,753)	29.681	17.581	29.896	29.273
	p-value ≤	.001	.001	.001	.001

Table 12. Means, Ns, standard deviations, and *p*-values for prominent and non-prominent words for the four Normalized LF measurements, pooled across speakers: EE (spectral intensity), RK (glottal symmetry), OQ (open quotient) and Linearity (spectral linearity).

These combined subjects results can be understood by reviewing the results from the individual subjects, summarized in Figure 43 and Table 13. Prominent words have tenser values than non-prominent words for at least one speaker for all four measurements and there are no significant results in the opposite direction. All three speakers show significantly greater spectral intensity (EE) for prominent words. One speaker, B, shows a significant effect of prominence on OQ, indicating that her prominent words have longer glottal closing durations (smaller values of OQ). B also shows a strong effect of prominence on spectral linearity (Lin); her prominent words have smaller spectral slopes

and/or fewer spectral notches than her non-prominent words. Speakers L and S showed a similar but non-significant effect. One speaker, S, shows an increased amounts of glottal skew (smaller values for RK) for prominent words. B shows a similar, but non-significant, effect.

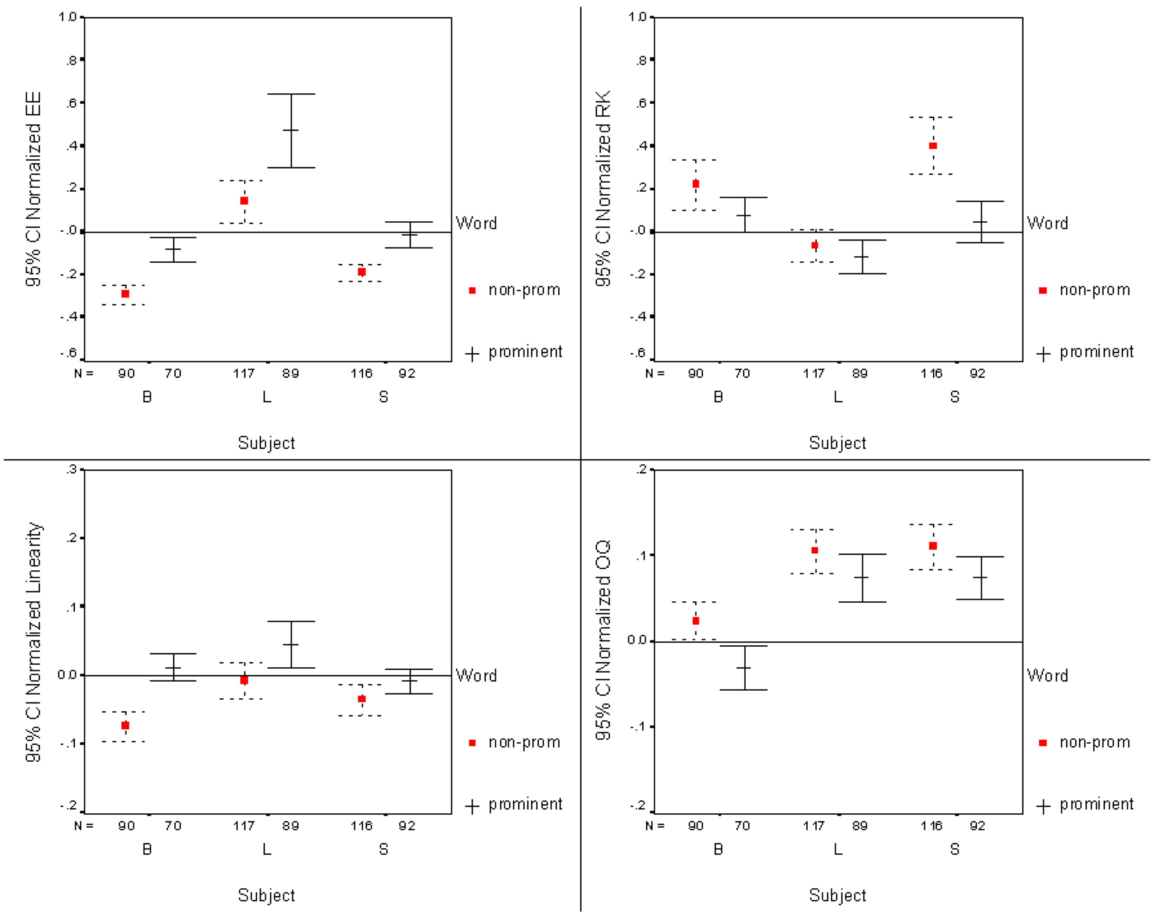


Figure 43 Means and 95% confidence interval error bars for prominent and non-prominent words for the four Normalized LF measurements: EE (spectral intensity), RK (glottal symmetry), OQ (open quotient) and Linearity (spectral linearity).

Prominent Words	Speaker B				Speaker L				Speaker S			
	EE	Lin	RK	OQ	EE	Lin	RK	OQ	EE	Lin	RK	OQ
Prominent Words Tenser	X	X		X	X				X		X	
Non-prom. Words Tenser												

Table 13. Summary of significant effects (solid 'X') of prominence on the voice quality of words.

5.2.2 Prominent vs. non-prominent stressed syllables

A comparison of syllable-medial samples from prominent and non-prominent stressed syllables found significant univariate effects for EE and OQ, and near significant effects for RK for the three speakers together. These results are summarized in Table 14. The MANOVA for the three speakers together found a significant main effect [F(1,316) = 3.962, *p* = 0.047]. There were, however, significant between-subject interactions for EE [F(2,312) = 4.960, *p* = 0.008] and Linearity [F(2,312) = 5.978, *p* = 0.003].

Stressed Syllable		Normalized EE	Normalized RK	Normalized OQ	Normalized Linearity
non-prom	Mean	-.15262	.19696	.10584	-.04732
	N	212	212	212	212
	Std. Deviation	.29478	.61211	.13694	.11734
prominent	Mean	-.08226	.07465	.04435	-.03113
	N	106	106	106	106
	Std. Deviation	.24998	.46187	.12368	.09078
	F(1,312)	7.043	3.732	14.146	2.412
	p-value ≤	.008	.054	.001	.121

Table 14. Means, Ns, standard deviations, and *p*-values for prominent and non-prominent syllables for the four Normalized LF measurements, pooled across speakers: EE (spectral intensity), RK (glottal symmetry), OQ (open quotient) and Linearity (spectral linearity).

These combined subjects results can be understood by reviewing the results from the individual subjects, summarized in Figure 44 and Table 15. Prominent syllables have tenser values than non-prominent syllables for either B or S for three of the four measurements; speaker L does not show any significant effects of prominent syllables on voice quality. Speaker S shows longer glottal closing durations (smaller values of OQ) for prominent syllables. Speakers B and L show similar non-significant trends. Speaker B shows a strong effect of prominence on spectral linearity (Linearity) and spectral intensity (EE), implying that her prominent syllables have both low values for spectral tilt

and greater intensity for all the harmonics. There are no significant effects of prominent syllables on glottal symmetry (RK) for any of the speakers. In short, the effect of prominent syllables on voice quality is strongest for one speaker, B – phrasally stressed syllables are “louder”; and this generally correlates with lower values of spectral tilt.

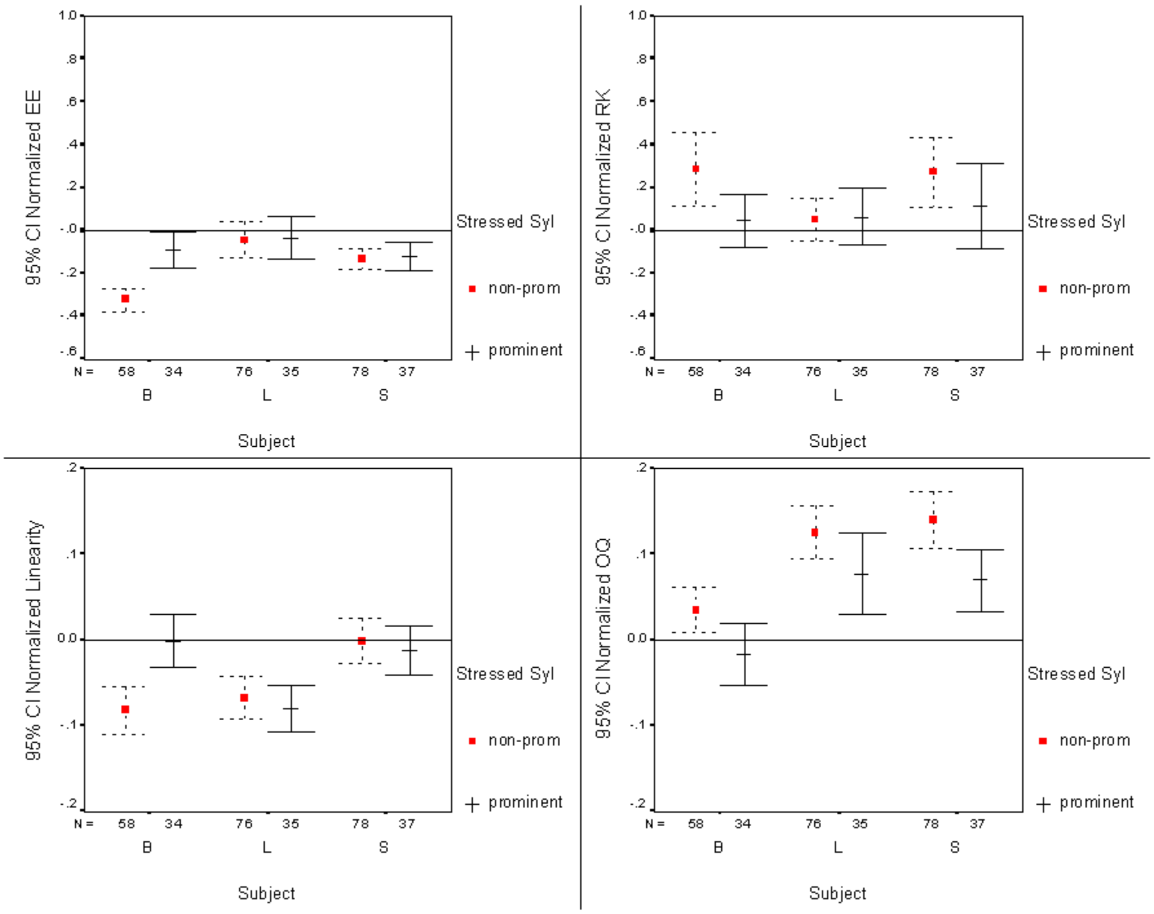


Figure 44. Means and 95% confidence interval error bars for prominent and non-prominent stressed syllables for the four Normalized LF measurements: EE (spectral intensity), RK (glottal symmetry), OQ (open quotient) and Linearity (spectral linearity).

Prominent Syllables	Speaker B				Speaker L				Speaker S			
	EE	Lin	RK	OQ	EE	Lin	RK	OQ	EE	Lin	RK	OQ
Prominent Syl. Tenser	X	X										X
Non-prom. Syl. Tenser												

Table 15. Summary of significant effects (solid 'X') of prominence on the voice quality of stressed syllables.

5.2.3 Prominent vs. non-prominent words in declaratives

A comparison of syllable-medial samples from prominent and non-prominent words in declarative sentences (this is the declarative subset of the results discussed in section 5.2.1) found significant univariate effects for all parameters except OQ for the three speakers combined (see Table 16). The MANOVA for the three speakers together found a significant main effect [$F(1,272) = 8.047$, $p \leq 0.005$]. There were, however, significant between-subject interactions for EE [$F(2,268) = 8.551$, $p \leq 0.001$], RK [$F(2,268) = 4.083$, $p \leq 0.018$] and Linearity [$F(2,268) = 7.913$, $p \leq 0.001$].

Words in Declarative Sentences		Normalized EE	Normalized RK	Normalized OQ	Normalized Linearity
non-prom	Mean	-.12211	.10516	.04761	-.03757
	N	143	143	143	143
	Std. Deviation	.33334	.46164	.12849	.11497
prominent	Mean	.15500	-.02622	.05640	.01759
	N	131	131	131	131
	Std. Deviation	.59265	.37215	.12237	.11560
	F(1,268)	23.974	4.420	.295	15.243
	p-value ≤	.001	.036	.588	.001

Table 16. Means, Ns, standard deviations, and *p*-values for prominent and non-prominent words in declarative sentences for the four Normalized LF measurements, pooled across speakers: EE (spectral intensity), RK (glottal symmetry), OQ (open quotient) and Linearity (spectral linearity).

These combined subjects results can be understood by reviewing the results from the individual subjects, summarized in Figure 45 and Table 16. Prominent words in declarative sentences have significantly tenser values than non-prominent words for Speaker L for three of the four measurements. B and S show similar, but non-significant trends. L shows greater spectral intensity (EE), greater glottal skew (smaller values of RK), and greater spectral linearity (Linearity). There are no effects of prominent words in declarative sentences on open quotient (OQ) for any of the speakers.

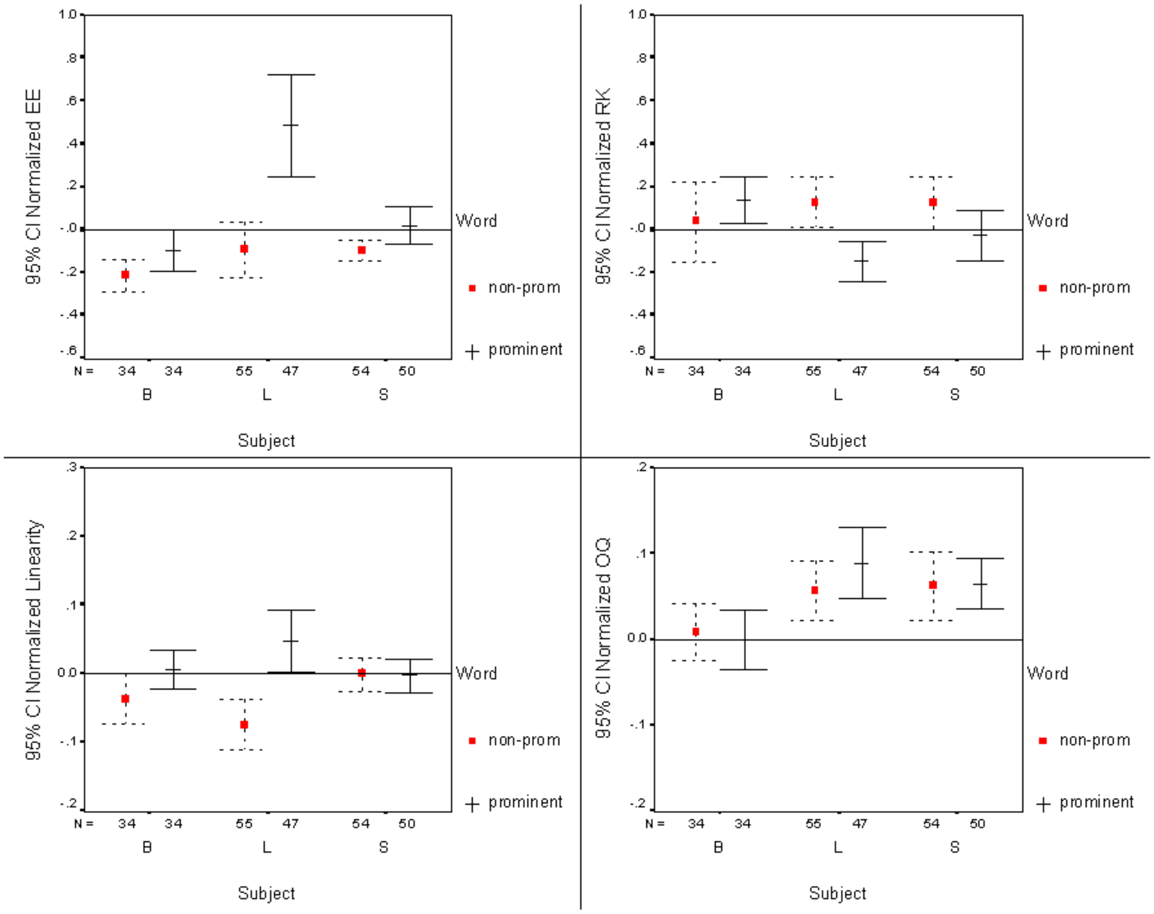


Figure 45. Means and 95% confidence interval error bars for prominent and non-prominent words in declarative sentences for the four Normalized LF measurements: EE (spectral intensity), RK (glottal symmetry), OQ (open quotient) and Linearity (spectral linearity).

Prominent Words in Declaratives	Speaker B				Speaker L				Speaker S			
	EE	Lin	RK	OQ	EE	Lin	RK	OQ	EE	Lin	RK	OQ
Prominent WordTenser					X	X	X		X (dashed)			
Non-prom. Word Tenser												

Table 17. Summary of significant effects (solid 'X') and nearly significant effects (dashed 'X') of prominence on the voice quality of words in declarative sentences.

5.2.4 Prominent vs. non-prominent words in interrogatives

A comparison of syllable-medial samples from prominent and non-prominent words in interrogative sentences (this is the interrogative subset of the results discussed in section 5.2.1) found significant univariate effects for all four parameters for the three speakers combined (see Table 18). The MANOVA for the three speakers together found a significant main effect [F(1,298) = 1.865, $p \leq 0.173$]; there were also significant between-subject interactions for RK [F(2,294) = 9.923, $p \leq 0.001$] and Linearity [F(2,294) = 5.621, $p \leq 0.004$].

Words in Interrogative Sentences		Normalized EE	Normalized RK	Normalized OQ	Normalized Linearity
non-prom	Mean	-.07992	.24322	.11393	-.03396
	N	180	180	180	180
	Std. Deviation	.47685	.70660	.13246	.13962
prominent	Mean	.12753	.02686	.03320	.01651
	N	120	120	120	120
	Std. Deviation	.58010	.43336	.13394	.12041
	F(1,294)	16.086	11.751	33.411	13.146
	p-value ≤	.001	.001	.001	.001

Table 18. Means, Ns, standard deviations, and *p*-values for prominent and non-prominent words in interrogative sentences for the four Normalized LF measurements, pooled across speakers: EE (spectral intensity), RK (glottal symmetry), OQ (open quotient) and Linearity (spectral linearity).

These combined subjects results can be understood by reviewing the results from the individual subjects, summarized in Figure 46 and Table 18. Prominent words in interrogatives have tenser values than non-prominent words for at least one speaker for all four measurements. Speakers B and L show longer glottal closing/closed durations (smaller values of OQ) for prominent words in interrogative sentences. Speaker S shows a nearly significant effect in the same direction. Two speakers, B and S, show strong effects of prominence for spectral intensity (EE). B and S also show increased amounts

of glottal skew (smaller values for RK) for prominent words. One speaker, B, shows a significant effect of prominence on spectral linearity (Linearity).

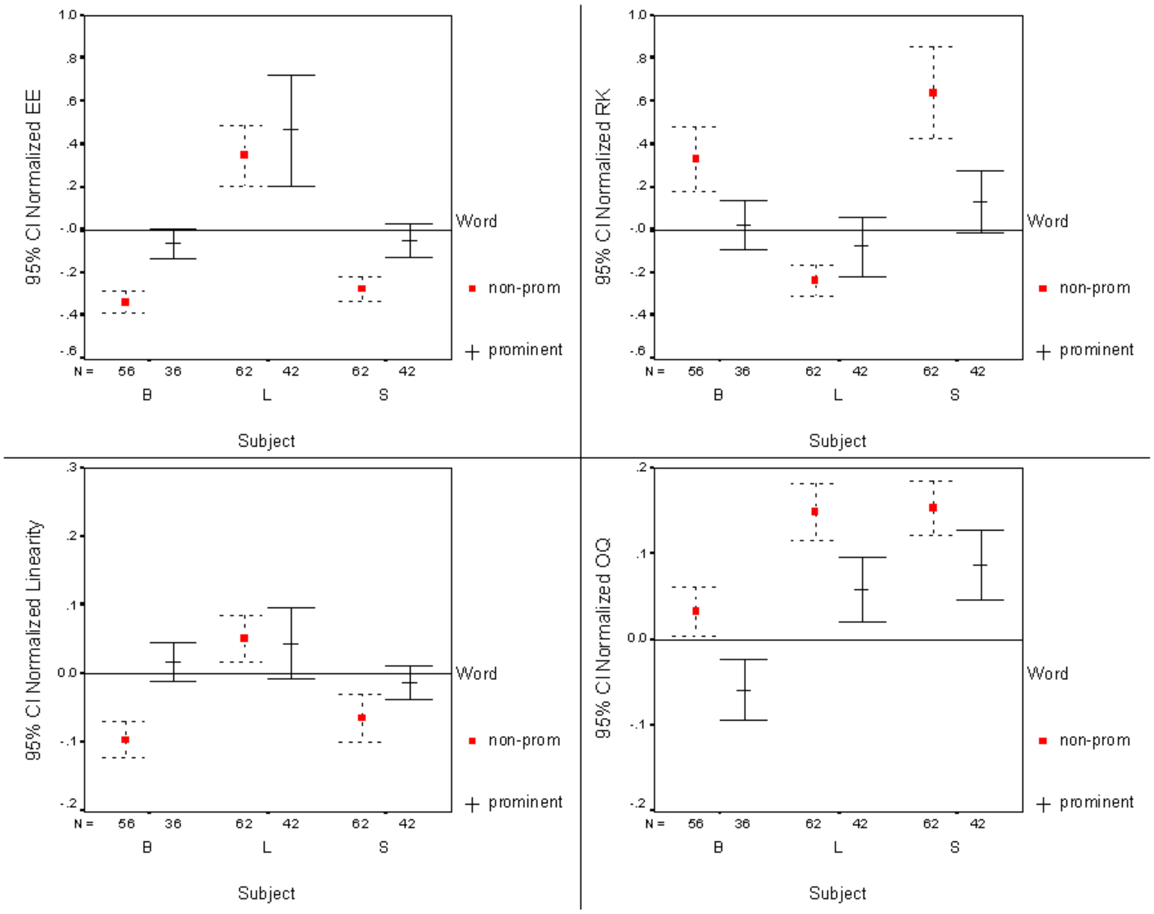


Figure 46. Means and 95% confidence interval error bars for prominent and non-prominent words in interrogative sentences for the four Normalized LF measurements: EE (spectral intensity), RK (glottal symmetry), OQ (open quotient) and Linearity (spectral linearity).

Prominent Words in Interrogatives	B				L				S			
	EE	Lin	RK	OQ	EE	Lin	RK	OQ	EE	Lin	RK	OQ
Prom Tenser	X	X	X	X		X (dashed)		X	X	X (dashed)	X	X (dashed)
Not Tenser												

Table 19. Summary of significant effects (solid 'X') and near significant effects (dashed 'X') of prominence on the voice quality of words in interrogative sentences.

5.2.5 Effects of prominence on non-modal phonation

Several studies have found increased amounts of allophonic word-initial vowel glottalization on prominent words (Pierrehumbert 1995; Dilley et al. 1996). Since prominent words in this corpus were found to have a tenser voice quality, this indicates that vowel-initial glottalization is associated with a tense voice quality. Although this corpus purposely did not contain any vowel-initial words, the proportion of non-LF-fitable waveforms that occurred in prominent and non-prominent words can be compared. As can be seen in Figure 47, these "non-modal" waveforms occurred most frequently in non-prominent words. Since non-prominent words were found to have laxer LF parameters during modal phonation, this indicates that non-vowel-initial creaky waveforms are associated with a lax voice quality.

In summary, there is a difference in the distribution of allophonic vowel-initial creak and syllable-medial non-modal phonation. Although the data for these results is acoustic, this suggests the possibility of different physiological causes for these two types of "creak".

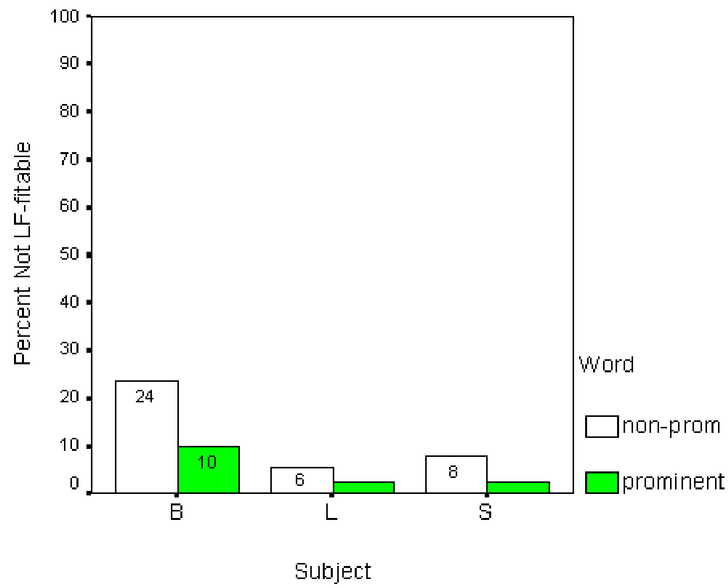


Figure 47. Percentage of not LF-fitable waveforms for prominent and non-prominent words for each speaker.

5.2.6 Discussion of the effects of prominence on voice quality

As can be seen from the results presented above, all three speakers show a strong effect of prominence on voice quality for prominent words, in that prominent words have a tenser voice quality than non-prominent words. This effect holds to a much smaller extent for prominent syllables, indicating that prominent voice quality is not a property of the stressed syllable, but a property of the word as a whole. Significant univariate effects over the entire group for many of the parameters indicate that the three speakers show the same trends, even if the effects for each individual may not be significant. As can be seen in Table 20, one speaker, L, shows the greatest effect for prominence in declarative sentences and the two female speakers, B and S, show the greatest effect for prominence in interrogatives. However, it is not clear why speakers should show a greater effect for prominence in only one sentence type. It is possible that this may be related to the different prominent tones in interrogative and declarative sentences, but as will be seen in section 5.4 on phonological tone, it is not clearly this either.

On the other hand, all three subjects use the parameter OQ, open quotient, only in the prominence distinction within interrogative sentences. In fact, one of the subjects, L does not show overlap at all in the parameter values used in declaratives and interrogatives. This indicates that longer vocal fold closure may be a distinctive characteristic of low prominent tones, which were used exclusively in interrogatives sentences in this corpus (although interrogatives did have high prominent tones 27.5% of the time). This will be discussed further in Chapter 6.

In summary, speakers do use voice quality to distinguish between prominent and non-prominent words. Individual speakers, though, may vary on whether this effect is stronger for prominent words in interrogative or declarative sentences.

	Speaker B				Speaker L				Speaker S			
	EE	Lin	RK	OQ	EE	Lin	RK	OQ	EE	Lin	RK	OQ
All Sentence Types	X	X		X	X				X		X	
Declaratives					X	X	X		X (dashed)			
Interrogatives	X	X	X	X		X (dashed)		X	X	X (dashed)	X	X (dashed)

Table 20. Parameters used to indicate greater tenseness for prominent words separated by sentence type and speaker. Significant effects are represented by a solid 'X' and near significant effects by a dashed 'X'. The parameters are: EE (spectral intensity), RK (glottal symmetry), OQ (open quotient) and Linearity (spectral linearity).

5.3 Phrase edges

Results of previous studies have been inconclusive regarding the effects of phrase edges on voice quality. Studies have found that there is an increase in creaky waveforms at both phrase-initial and phrase-final boundaries (Hagen 1997; Redi & Shattuck-Hufnagel 2001). Studies have also found both a tenser and a laxer voice quality at the end of the sentence (Klatt & Klatt 1990; Gobl 1988).

As noted earlier, there are a number of ways the effects of phrase-boundaries can be studied: on a microscopic level, by comparing each syllable/word of a sentence; on a macroscopic level, by comparing large sections of a sentence with each other; or at a middle level, by comparing phrase-initial and phrase-final words with each other. Since the test corpus is designed so that prominent words occur at the phrase edges, this study will primarily examine the effects of phrase edges on the voice quality of prominent words located at phrase-final and phrase-initial edges, both for all sentence types together and for interrogatives and declaratives separately. It will also examine the effects of phrase edges on large sections of the sentence. As can be seen from the studies discussed earlier, it is not clear what these effects should be, since the results of previous studies are inconsistent and are primarily from declarative sentences. Thus, the following hypotheses will be tested:

- **Phrase-initial prominent words have a different voice quality than phrase-final prominent words for all sentence intonations combined.**
- **Phrase initial sections of the sentence have a different voice quality than phrase-final sections of the sentence for all sentence intonations combined.**
- **Phrase-initial prominent words have a different voice quality than phrase-final prominent words for declarative sentences.**
- **Phrase-initial prominent words have a different voice quality than phrase-final prominent words for interrogative sentences.**

5.3.1 Prominent words at phrase edges

A comparison of phrase-initial and phrase-final prominent words found significant univariate effects for all four LF measurements for the three speakers combined. These results are summarized in Table 21.[6] The MANOVA for the three speakers together found a significant main effect [F(1,434) = 9.765, $p \leq 0.002$]. There were, however, significant between-subject interactions for EE [F(2,430) = 8.156, $p \leq 0.001$], RK [F(2,430) = 15.169, $p \leq 0.001$] and Lin [F(2,430) = 10.850, $p \leq 0.001$].

Sentence Position of Prominent Words		Normalized EE	Normalized RK	Normalized OQ	Normalized Linearity
Initial	Mean	.17649	-.12582	.01255	.02713
	N	258	258	258	258
	Std. Deviation	.53921	.31496	.12617	.10260
Final	Mean	-.10502	.24923	.06037	-.01909
	N	178	178	178	178
	Std. Deviation	.33428	.50698	.12552	.12028
	F(1,430)	39.807	91.995	23.495	19.741s
	p-value ≤	.001	.001	.001	.001

Table 21. Means, Ns, standard deviations, and *p*-values for prominent words at prosodic edges for all sentence types combined for the four Normalized LF measurements, pooled across speakers: EE (spectral intensity), RK (glottal symmetry), OQ (open quotient) and Linearity (spectral linearity).

These combined subjects results can be understood by reviewing the results from the individual subjects, summarized in Figure 48 and Table 22. Sentence-initial prominent words have tenser values for at least one speaker for all four LF measurements. No speakers show significant effects in the opposite direction. Two speakers, L and S, show increased amounts of glottal skew (smaller values of RK) for sentence-initial prominent words. B has a non-significant effect in the same direction. Two speakers, B and L, show a strong effect of sentence position on glottal closing durations, in that the

[6] Each stressed syllable was represented by syllable-initial, -medial and -final samples. Each unstressed syllable was represented by a syllable-medial sample.

sentence-initial tokens have longer closing durations (smaller values of OQ). Two speakers, L and S, show a strong effect of sentence position on spectral intensity (EE) – their sentence-initial words having greater intensity than their sentence-final words. The same two speakers show a strong effect of sentence position on spectral linearity (Linearity); their sentence-initial prominent words had smaller spectral slopes and/or fewer spectral notches than their sentence final words.

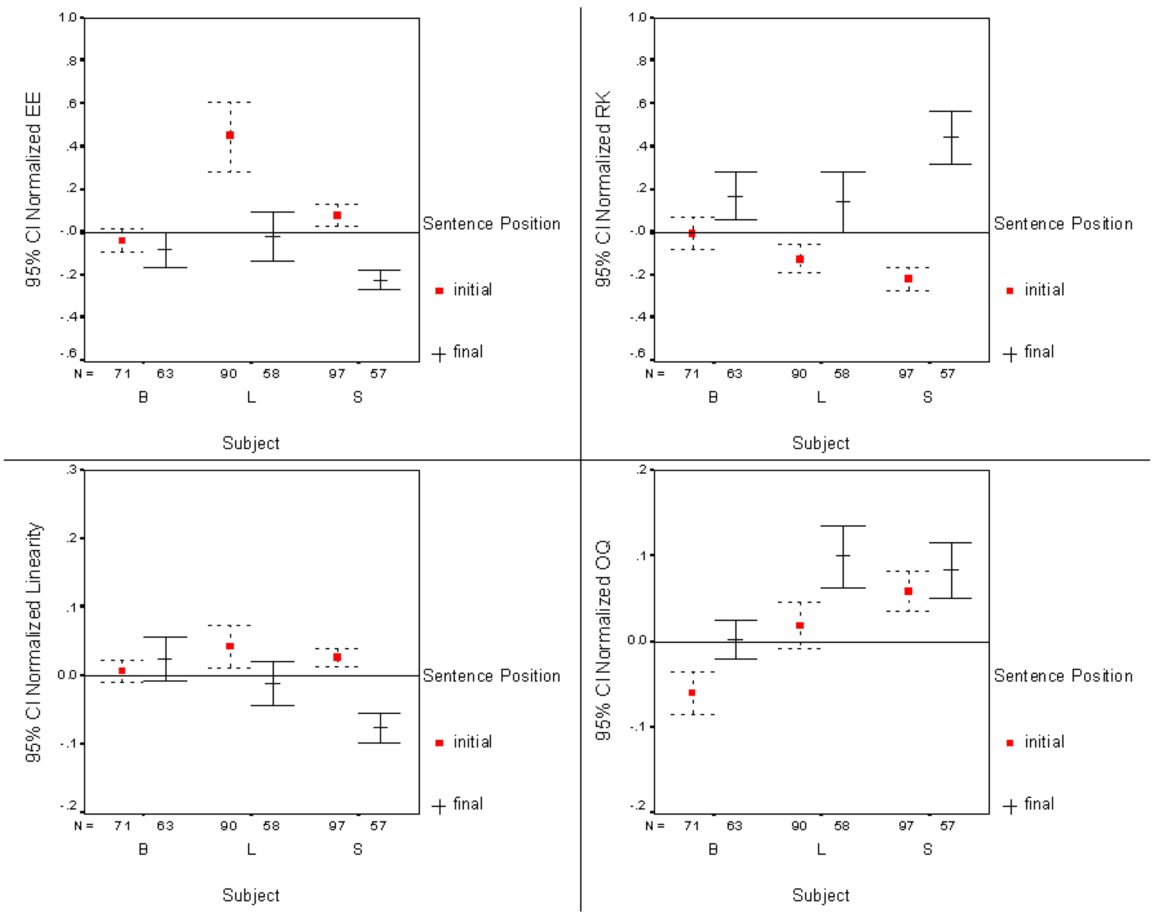


Figure 48. Means and 95% confidence interval error bars for prominent words at prosodic edges for all sentence types for the four Normalized LF measurements: EE (spectral intensity), RK (glottal symmetry), OQ (open quotient) and Linearity (spectral linearity).

Prominent Words at Phrase Edges:	Speaker B				Speaker L				Speaker S			
All Sent. Types	EE	Lin	RK	OQ	EE	Lin	RK	OQ	EE	Lin	RK	OQ
Initial Tenser			X (thin)	X	X	X (dashed)	X	X	X	X	X	
Final Tenser												

Table 22. Summary of significant effects (solid 'X') and nearly significant effects (dashed 'X') of phrase edge on the voice quality of prominent words.

5.3.2 Large sections at phrase edges

In this test, phrase-initial (*Dagada gave*) and phrase-final (*Bobby doodads*) sections of the sentence *Dagada gave Bobby doodads* for all sentence types were compared. Significant univariate effects for all four LF measurements for the three speakers combined were found. These results are summarized in Table 23.[7] The MANOVA for the three speakers together did not find a significant main effect [F(1,1033) = 0.004, $p \leq$ 0.951]. There were also significant between-subject interactions for all four parameters: EE [F(2,1029) = 10.954, $p \leq 0.001$], RK [F(2,1029) = 27.717, $p \leq 0.001$], OQ [F(2,1029) = 13.264, $p \leq 0.001$] and Linearity [F(2,1029) = 10.008, $p \leq 0.001$].

Sentence Position		Normalized EE	Normalized RK	Normalized OQ	Normalized Linearity
Initial	Mean	.11989	-.01054	.02164	.01136
	N	457	457	457	457
	Std. Deviation	.53332	.44801	.13044	.11071
Final	Mean	-.08901	.14230	.07960	-.02056
	N	578	578	578	578
	Std. Deviation	.41374	.52873	.13079	.12730
	F(1,1029)	59.040	26.166	52.915	18.205
	p-value ≤	.001	.001	.001	.001

Table 23. Means, Ns, standard deviations, and *p*-values for initial and final sections of all sentence types for the four Normalized LF measurements, pooled across speakers: EE (spectral intensity), RK (glottal symmetry), OQ (open quotient) and Linearity (spectral linearity).

These combined subject results can be understood by reviewing the results from the individual subjects, summarized in Figure 49 and Table 24. Sentence-initial sections have tenser values for all four parameters for at least one speaker, and there are no significant effects in the opposite direction. Speaker B shows no significant effects. Two

[7] Prominent stressed syllables and sentence-final syllables were represented by syllable-initial, -medial and -final samples. Unstressed syllables and non-prominent stressed syllables were represented by syllable-medial samples.

of the speakers, L and S, show effects of sentence position for spectral intensity (large values of EE) and glottal closing durations (small values of OQ). L also has significant effects for spectral linearity (large values of Linearity) and glottal skew (small values of RK). These results, while similar to those of the prominent words, are not nearly as strong, particularly for speaker B. This suggests that phrase edge effects may be strongest for either prominent words or individual words, although longer, more complex sentences may give a better indication of the extent of the effects of phrase edges.

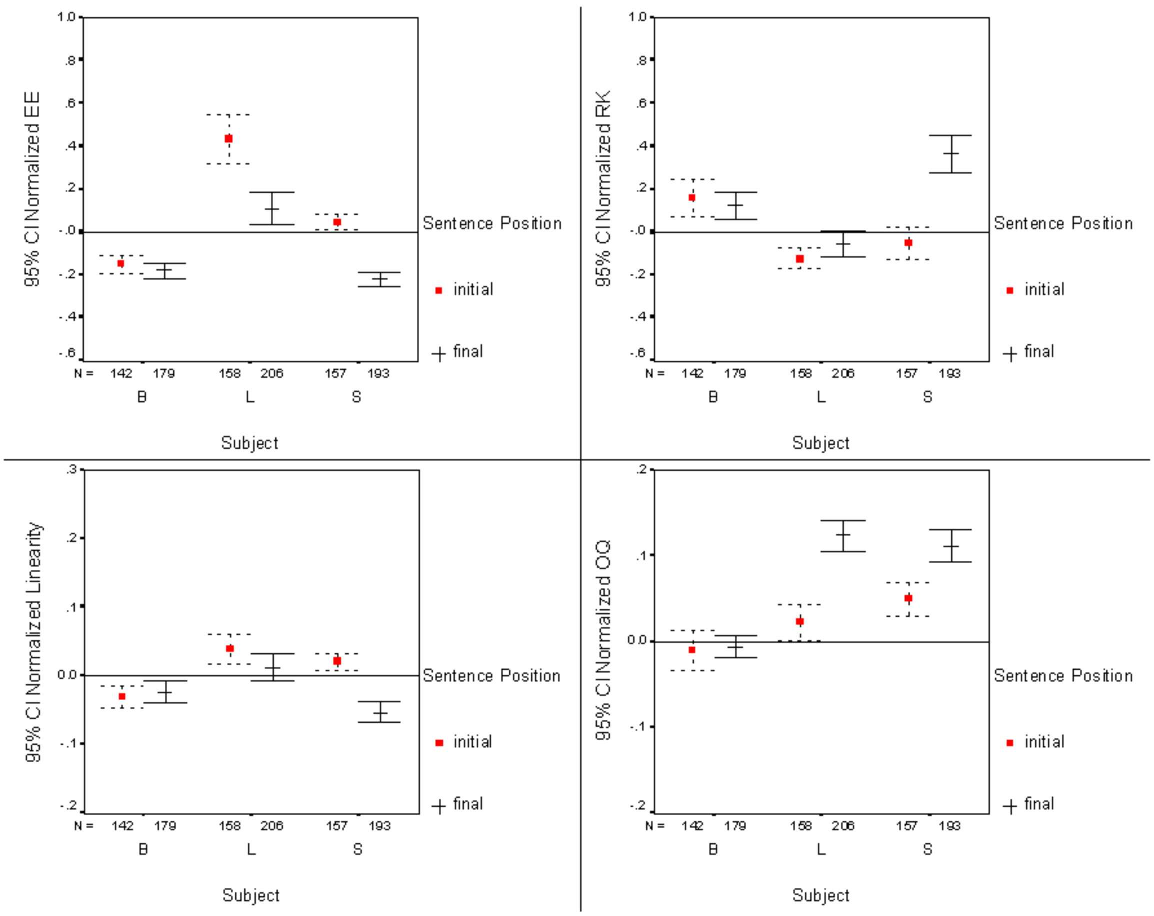


Figure 49. Means and 95% confidence interval error bars for initial and final sections for all sentence types for the four Normalized LF measurements: EE (spectral intensity), RK (glottal symmetry), OQ (open quotient) and Linearity (spectral linearity).

Large Sections at Phrase Edges	Speaker B				Speaker L				Speaker S			
All Sentence Types	EE	Lin	RK	OQ	EE	Lin	RK	OQ	EE	Lin	RK	OQ
Initial Tenser					X			X	X	X	X	X
Final Tenser												

Table 24. Summary of significant effects (solid 'X') of phrase edges on the voice quality of large sections of the sentence. All sentence types are combined.

5.3.3 Prominent words at the edges of declarative sentences

A comparison of phrase-initial and phrase-final prominent words in declarative sentences found significant univariate effects for all four parameters for the three speakers combined (this is the declarative subset of the results discussed in section 5.3.1). These results are summarized in Table 25.[8] The MANOVA for the three speakers together found a significant main effect [$F(1,221) = 2.136$, $p \leq 0.145$]. There were, however, significant between-subject interactions for all four parameters: EE [$F(2,217) = 8.711$, $p \leq 0.001$], RK [$F(2,217) = 8.435$, $p \leq 0.001$], OQ [$F(2,217) = 3.919$, $p \leq 0.021$] and Lin [$F(2,217) = 5.049$, $p \leq 0.007$].

Sentence Position of Prominent Words in Declaratives		Normalized EE	Normalized RK	Normalized OQ	Normalized Linearity
Initial	Mean	.20463	-.16125	.03316	.04066
	N	139	139	139	139
	Std. Deviation	.54618	.30142	.12948	.09680
Final	Mean	-.13594	.22728	.07895	-.04015
	N	84	84	84	84
	Std. Deviation	.27209	.45312	.11990	.10606
	F(1,217)	30.380	60.735	11.230	36.591
	p-value ≤	.001	.001	.001	.001

Table 25. Means, Ns, standard deviations, and *p*-values for prominent words at phrase edges of declaratives for all sentence types for the four Normalized LF measurements, pooled across speakers: EE (spectral intensity), RK (glottal symmetry), OQ (open quotient) and Linearity (spectral linearity).

These combined subjects results can be understood by reviewing the results from the individual subjects, summarized in Figure 50 and Table 26. Sentence-initial prominent words in declaratives have tenser values for at least one speaker for all four LF parameters. Two speakers, S and L, show increased amounts of glottal skew (smaller

[8] Each stressed syllable was represented by syllable-initial, -medial and -final samples. Each unstressed syllable was represented by a syllable-medial sample.

values of RK) for sentence-initial prominent words. They also show a significant effect of sentence position on spectral intensity (EE) and spectral linearity (Linearity). Their sentence-initial words have greater overall spectral intensity as well as smaller spectral slopes and/or fewer spectral notches. Speaker B showed an effect of sentence position on open quotient (OQ). Her sentence initial words had a longer closing duration (smaller OQ) than her sentence-final words. Speaker L shows a similar, but non-significant effect. In summary, two speakers, L and S, are using the same set of parameters to distinguish between phrase-initial and phrase-final prominent words in declaratives. The third speaker, shows the same effect, but uses a different parameter.

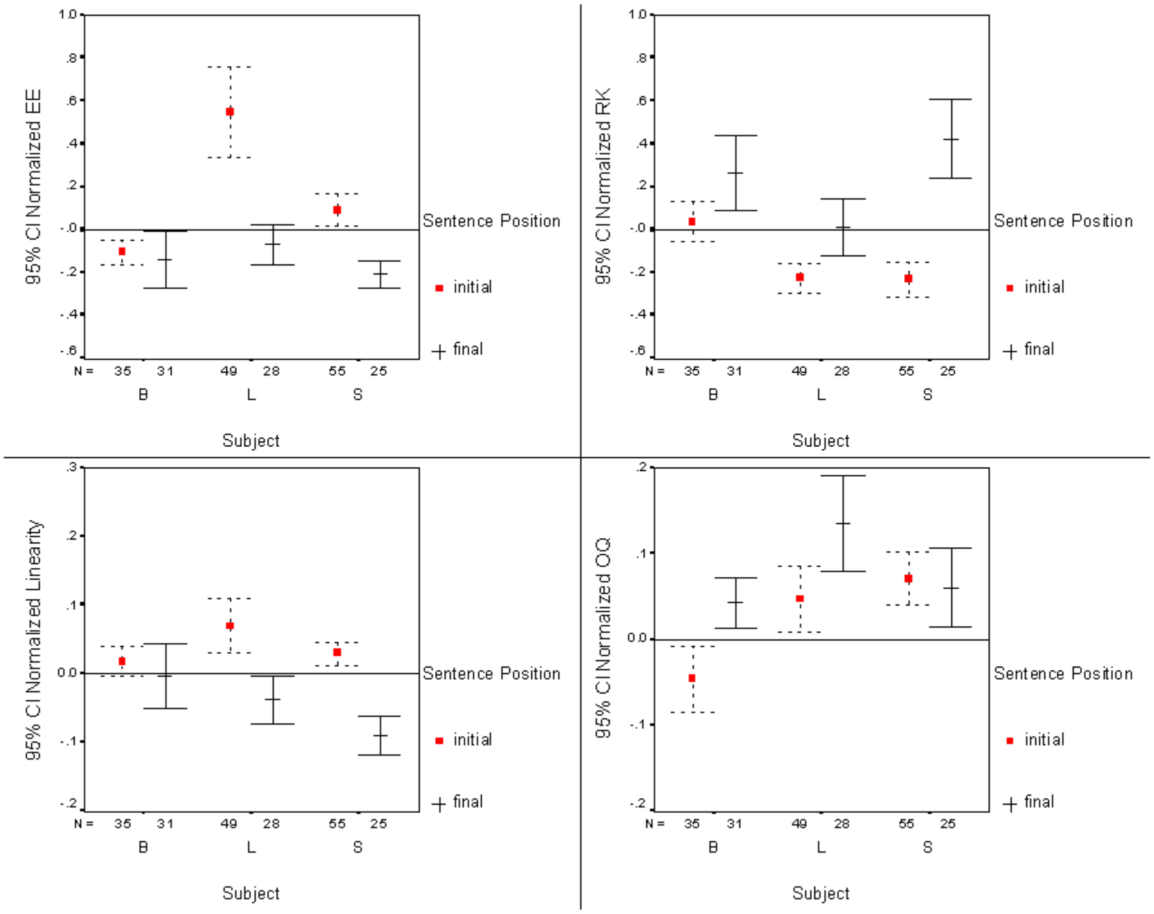


Figure 50. Means and 95% confidence interval error bars for prominent words at phrase edges of declaratives for the four Normalized LF measurements: EE (spectral intensity), RK (glottal symmetry), OQ (open quotient) and Linearity (spectral linearity).

Prominent Words at Phrase Edges:	Speaker B				Speaker L				Speaker S			
Declaratives	EE	Lin	RK	OQ	EE	Lin	RK	OQ	EE	Lin	RK	OQ
Initial Tenser				X	X	X	X	(X)	X	X	X	
Final Tenser												

Table 26. Summary of significant effects (solid 'X') and nearly significant effects (dashed 'X') of phrase edge on the voice quality of prominent words in declaratives.

5.3.4 Prominent words at phrase edges of interrogatives

A comparison of phrase-initial and phrase-final prominent words in interrogative sentences found significant univariate effects for three of the four LF measurements: EE, RK and OQ (this is the interrogative subset of the results discussed in section 5.3.1). These results are summarized in Table 27.[9] The MANOVA for the three speakers together found a significant main effect [F(1,211) = 8.529, $p \leq 0.004$]. There were, however, significant between-subject interactions for two parameters: RK [F(2,207) = 7.542, $p \leq 0.001$] and Linearity [F(2,207) = 6.946, $p \leq 0.001$].

Sentence Position		Normalized EE	Normalized RK	Normalized OQ	Normalized Linearity
Initial	Mean	.14362	-.08444	-.01152	.01133
	N	119	119	119	119
	Std. Deviation	.53135	.32649	.11823	.10724
Final	Mean	-.07740	.26884	.04378	-.00027
	N	94	94	94	94
	Std. Deviation	.38074	.55237	.12870	.12936
	F(1,207)	11.499	35.333	14.337	.457
	p-value ≤	.001	.001	.001	.500

Table 27. Means, Ns, standard deviations, and *p*-values for prominent words at prosodic edges of interrogatives for all sentence types for the four Normalized LF measurements, pooled across speakers: EE (spectral intensity), RK (glottal symmetry), OQ (open quotient) and Linearity (spectral linearity).

These combined subjects results can be understood by reviewing the results from the individual subjects, summarized in Figure 51 and Table 28. Sentence-initial prominent words have tenser values for one speaker, S, for three of the four LF parameters. Speakers L and B show similar, but non-significant trends and none of the speakers show significant trends in the opposite direction. Speaker S's sentence-initial prominent words are characterized by high levels of spectral intensity (EE) and spectral linearity

[9] Each stressed syllable was represented by syllable-initial, -medial and -final samples. Each unstressed syllable was represented by a syllable-medial sample.

(Linearity). Her glottal pulses are characterized by large values of glottal skew (small values of RK). Speaker L has nearly significant sentence-initial effects for open quotient (OQ) – his glottal pulses are characterized by longer closing durations (small values of OQ). Speaker B, on the other hand, does not show any significant effects of sentence position in interrogatives. Interestingly, although none of the subjects show a significant effect of sentence position on OQ, there is a significant univariate effect (see Table 27), indicating that the subjects' overall trend is strong enough for them to have an effect together. In summary, although only one speaker shows significant effects of phrase edges on the voice quality of prominent words in interrogatives, there is a strong enough trend among the other speakers to cause significant univariate effects.

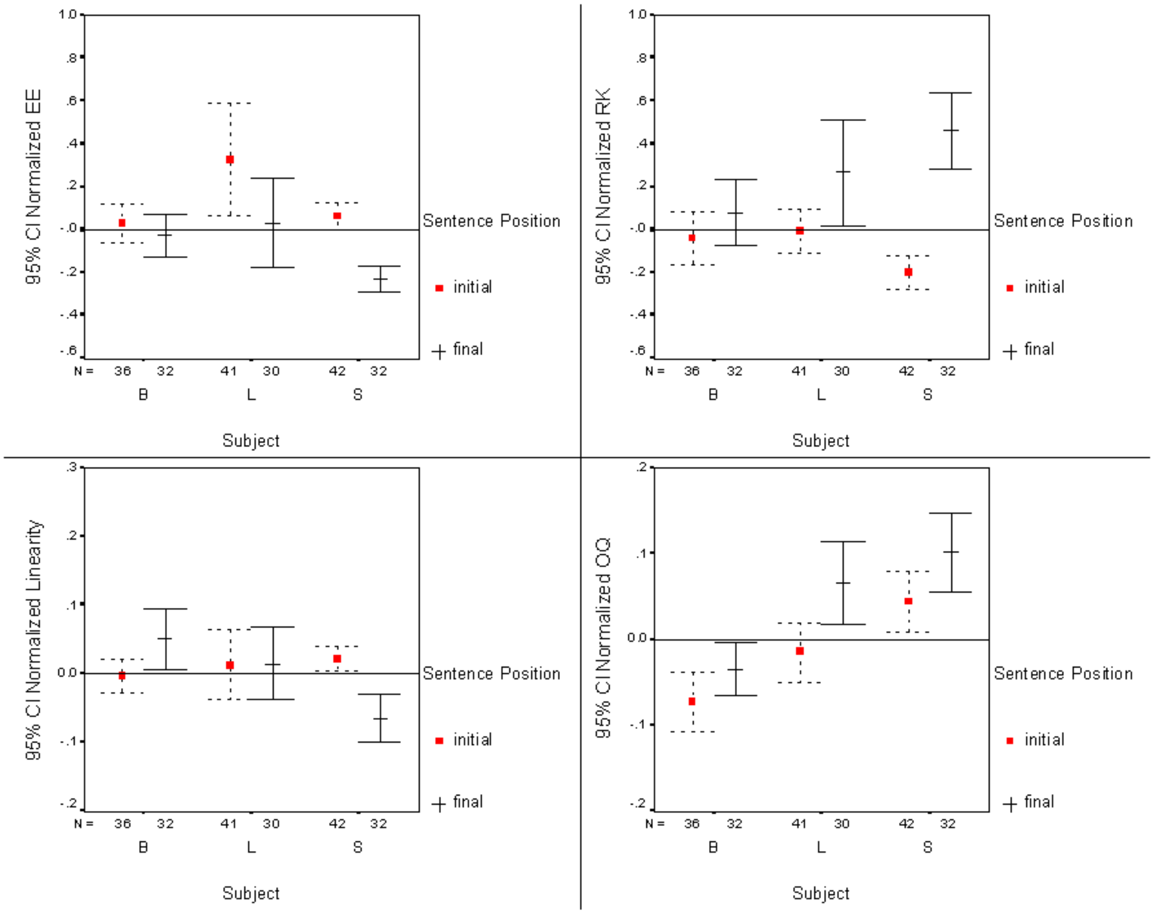


Figure 51. Means and 95% confidence interval error bars for prominent words at phrase edges of interrogatives for the four Normalized LF measurements: EE (spectral intensity), RK (glottal symmetry), OQ (open quotient) and Linearity (spectral linearity).

Prominent Words at Phrase Edges:	Speaker B				Speaker L				Speaker S			
Interrogatives	EE	Lin	RK	OQ	EE	Lin	RK	OQ	EE	Lin	RK	OQ
Initial Tenser								X (dashed)	X	X	X	
Final Tenser												

Table 28. Summary of significant effects (solid 'X') and nearly significant effects (dashed 'X') of phrase edge on the voice quality of prominent words in interrogatives.

5.3.5 Effects of phrase edges on non-modal phonation

Several studies have found increased amounts of allophonic vowel-initial glottalization and creaky waveforms at both phrase-initial and phrase-final boundaries (Pierrehumbert & Talkin 1992; Dilley et al. 1996; Hagen 1997; Redi & Shattuck-Hufnagel 2001). Results from this study, looking at all sentence types together as these studies have done, support these findings.

As can be seen in Figure 52, two speakers, B and S, produce non-modal phonation on the phrase-initial first syllable of *Dagada*. Since in the base corpus none of the speakers produced non-modal phonation on this syllable, the effect of phrase-initial position on the proportion of non-modal phonation is significant for speaker B (chi-square = 4480.513, $p \leq 0.001$) and speaker S (chi-square = 275.558, $p \leq 0.001$).

Furthermore, all three subjects produced non-modal phonation on the phrase-final last syllable of *doodads*. Two of the speakers produced significantly more non-modal phonation on *dads* in the test corpus than in the base corpus: speaker L (chi-square = 9.544, $p \leq 0.002$) and speaker S (chi-square = 14.816, $p \leq 0.001$). Speaker B, on the other hand, produced a greater proportion of non-modal phonation in the base corpus than in the test corpus (chi-square = 10.490, $p \leq 0.001$), suggesting that the non-modal phonation that appears on *dads* could be caused by her pronunciation tendencies. However, speaker B also produced the greatest proportion of non-modal phonation in both the test and base corpora, as seen in Table 29 and Figure 52, indicating that she does have an overall "creakier" voice. Since the sentence-final syllable *dads* in the test corpus

does have the second-highest proportion of non-modal phonation, it is possible that this syllable is still significantly "creakier" for her.

In summary, the speakers as a group produced significantly more non-modal phonation on syllables that are in phrase-initial or phrase-final position than on syllables that are in the phrase-medial baseline position.

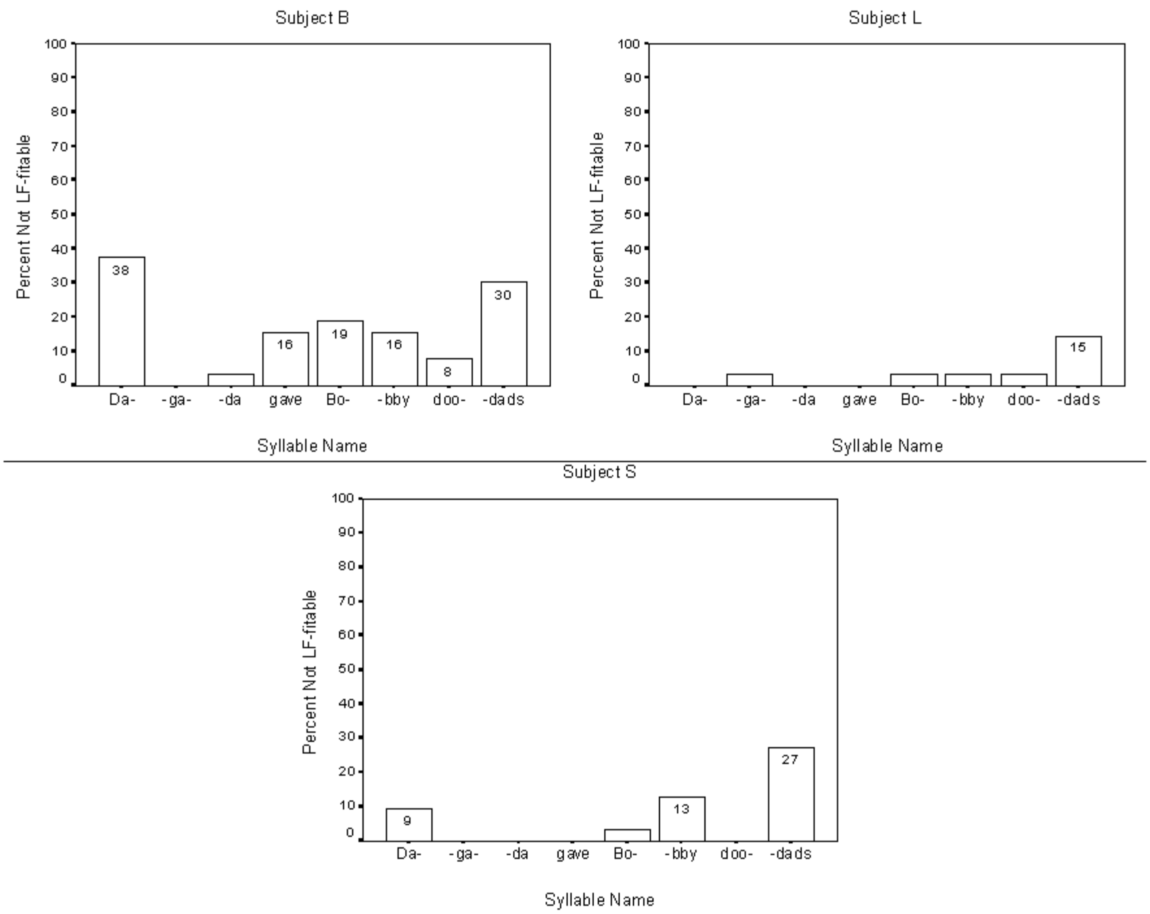


Figure 52. Percentage of non-LF-fitable waveforms for each syllable of the sentence for each speaker. All sentence types are combined.

	Percent Non-LF-Fitable Tokens								
	Base Syllable Name								Percentage of all base tokens
	Da-	-ga-	-da	gave	Bo-	-bby	doo-	-dads	
Speaker B			20.0%			40.0%	6.7%	46.7%	13.8%
Speaker L			40.0%					6.7%	3.8%
Speaker S					6.7%			13.3%	3.8%

Table 29. Percentage of non-LF-fitable tokens in the base corpus for each speaker.

5.3.6 Discussion of the effects of phrase edges on voice quality

The results presented above show a strong effect of phrase edges on the voice quality of prominent words (see Table 30). When looking at all sentence types together, phrase-initial prominent words have a tenser voice quality than phrase-final prominent words. This effect holds for two of the three speakers for large sections of the sentence, as well. For one speaker, S, this effect does not depend on the sentence intonation tune. Speakers B and L, on the other hand, show the strongest effects of phrase edges for declaratives, although neither of them show significant trends in the opposite direction for interrogatives. It is not clear why some speakers only show a significant effect of phrase edges for only one sentence type. Significant univariate effects over the entire group for many of the parameters, however, indicate that the three speakers do show the same general trends, even if their effects are not significant. Furthermore, all three speakers show increased proportions of non-modal phonation at phrase edges, as well.

	Speaker B				Speaker L				Speaker S			
	EE	Lin	RK	OQ	EE	Lin	RK	OQ	EE	Lin	RK	OQ
Phrase-initial tenser			X (dashed)	X	X	X (dashed)	X	X	X	X	X	
Declarative-initial tenser				X	X	X	X	X (dashed)	X	X	X	
Interrogative-initial tenser								X (dashed)	X	X	X	

Table 30. Parameters used to indicate greater tenseness for phrase edges separated by sentence type and speaker. Significant effects are represented by a solid 'X' and near significant effects by a dashed 'X'. The parameters are: EE (spectral intensity), RK (glottal symmetry), OQ (open quotient) and Linearity (spectral linearity).

In summary, data from this corpus suggests a "weakening" of voice quality across the sentence. However, it is not possible to tell from this data if the weakening is declination, initial strengthening or final weakening (c.f. Fougeron & Keating 1997). To

find out what kind of weakening is occurring, data would need to be collected from a corpus with more prominent word positions and more complex, nested prosodic structures.

5.4 Phonological tone

As discussed earlier, there is no a priori reason to suppose a relationship between phonetic pitch and voice quality. And in fact, for two out of three speakers in this corpus there is no correlation between F0 and voice quality. However, there is the possibility that a local phonological low or high, in other words a Low or High prominent tone or boundary tone, could cause a change in voice quality locally. As mentioned in section 4.6, High and Rise tones are collapsed into a single High tone category, and Low and Fall tones are collapsed into a single Low tone category. Boundary tones and prominent tones are, of course, kept separate in the analysis. The following hypotheses are tested in this section:

- **Syllables with High prominent tones have a tenser voice quality than syllables with Low prominent tones.**
- **Words with High prominent tones have a tenser voice quality than words with Low prominent tones.**
- **Syllables with High boundary tones have a tenser voice quality than syllables with Low boundary tones.**

5.4.1 Phonological tone of syllable

A comparison of three samples (syllable-initial, -medial and -final) from syllables with High and Low prominent tones did not find a consistent effect across parameters or across speakers. The effects for all three speakers combined are summarized in Table 31. There were significant univariate effects for all four LF measurements for the three speakers combined. However, three of the parameters, EE, RK and Linearity, indicated that High tones are tenser than Low prominent tones, and OQ indicated that Low tones are tenser than High tones. The MANOVA for the three speakers together did find a significant main effect [F(1,289) = 4.623, $p \leq 0.032$]. There were, however, significant between-subject interactions for EE [F(2,285) = 25.034, $p \leq 0.001$], RK [F(2,285) = 15.706, $p \leq 0.001$] and Lin [F(2,285) = 3.744, $p \leq 0.025$].

Syllable Tone		Normalized EE	Normalized RK	Normalized OQ	Normalized Linearity
High	Mean	-.03248	-.02356	.05146	-.00277
	N	183	183	183	183
	Std. Deviation	.24577	.45343	.12724	.08666
Low	Mean	-.10688	.22534	-.01948	-.03219
	N	108	108	108	108
	Std. Deviation	.28271	.48900	.10942	.11733
	F (1,285)	11.030	26.300	13.311	9.502
	p-value ≤	.001	.001	.001	.002

Table 31. Means, Ns, standard deviations, and *p*-values for tones on prominent syllables for the four Normalized LF measurements, pooled across speakers: EE (spectral intensity), RK (glottal symmetry), OQ (open quotient) and Linearity (spectral linearity).

A review of the individual subjects' results, summarized in Figure 53 and Table 32, helps explain why there are inconsistent effects for the speakers as a group. Only one parameter, OQ (open quotient), shows a consistent trend across the three speakers. This was also the only parameter to have both a significant univariate effect for the group and

no significant between-subject interactions. This indicates that it is likely that there is an effect across speakers of prominent pitch accent tone on OQ, in that Low tones have longer closing durations (smaller OQ) than high tones. On the other hand, speakers B and S generally show opposite effects of phonological tone on voice quality, and L uses two parameters to indicate that High tones are tenser than Low tones (EE (spectral intensity) and RK (glottal symmetry)) and one parameter (OQ) to indicate that Low tones are tenser than High tones. Furthermore, the parameter EE shows that High tones are tenser than Low tones for speakers L and S, but shows the opposite effect for speaker B.

In summary, only one parameter, OQ, shows a consistent trend across speakers. This parameter indicates that Low tones have a tenser voice quality than High tones. Two speakers, however, show significant trends for High tones having a tenser voice quality than Low tones. In other words, there is a great deal of personal variation in the effects of phonological tone on voice quality. Individual speakers each have their own unique trends, and the one parameter that is consistent across speakers, does not tell the whole story.

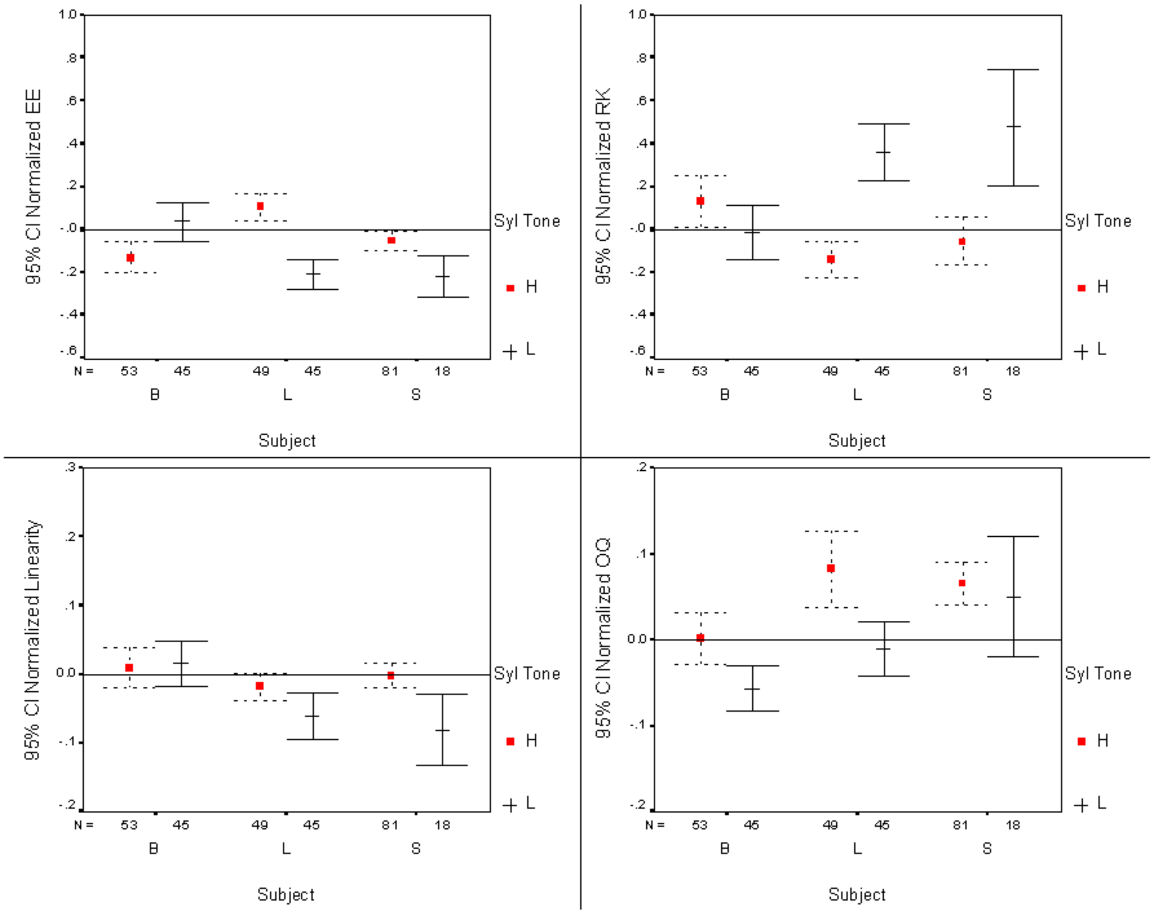


Figure 53. Means and 95% confidence interval error bars for tone of prominent syllables for the four Normalized LF measurements: EE (spectral intensity), RK (glottal symmetry), OQ (open quotient) and Linearity (spectral linearity).

Syllable Tone	Speaker B				Speaker L				Speaker S			
	EE	Lin	RK	OQ	EE	Lin	RK	OQ	EE	Lin	RK	OQ
H Tenser					X	X (dashed)	X		X	X	X	
L Tenser	X			X				X				

Table 32. Summary of significant effects (solid 'X') and nearly significant effects (dashed 'X') of phonological tone on the voice quality of prominent syllables.

5.4.2 Phonological tone of word

In case phonological tone, like prominence, had a more consistent effect on the voice quality of words instead of stressed syllables, a comparison of samples from words with High and Low prominent tones was also performed. It did not find a consistent effect across parameters or across speakers either. The results for the three speakers combined are summarized in Table 33.[10] There were significant univariate effects for three of the four LF measurements for the three speakers combined. However, two of the parameters, RK and Linearity, indicated that High tones are tenser than Low prominent tones, and OQ indicated the opposite. The MANOVA for the three speakers together did find a significant main effect [$F(1,434) = 2.593$, $p \leq 0.108$]. There were, however, significant between-subject interactions for EE [$F(2,430) = 5136$, $p \leq 0.006$], RK [$F(2,430) = 15.075$, $p \leq 0.001$] and OQ [$F(2,430) = 3.622$, $p \leq 0.028$].

Word Tone		Normalized EE	Normalized RK	Normalized OQ	Normalized Linearity
H	Mean	.06785	-.02405	.05073	.01106
	N	284	284	284	284
	Std. Deviation	.46108	.41259	.12657	.10236
L	Mean	.04981	.12324	-.00278	.00302
	N	152	152	152	152
	Std. Deviation	.53146	.48474	.12351	.12916
	F (1,430)	2.884	18.541	6.161	4.160
	p-value ≤	.090	.001	.013	.042

Table 33. Means, Ns, standard deviations, and *p*-values for tones on prominent words for the four Normalized LF measurements, pooled across speakers: EE (spectral intensity), RK (glottal symmetry), OQ (open quotient) and Linearity (spectral linearity).

A review of the individual subjects' results, summarized in Figure 54 and Table 34, helps explain why there are inconsistent effects for the speakers as a group. Unlike the

[10] Each stressed syllable was represented by syllable-initial, -medial and -final samples. Each unstressed syllable was represented by a syllable-medial sample.

results for prominent syllables, though, there is not a single measurement that shows consistent trends across speakers. Speakers B and S show opposite effects of phonological tone on the voice quality of prominent words, and L uses one parameter to indicate that High tones are tenser than Low tones (RK) and one parameter (OQ) to indicate that Low tones are tenser than High tones. Furthermore, as with the syllables, the parameter EE shows that High tones are tenser than Low tones for S, but shows the opposite effect for B.

In summary, like the results for prominent syllables, there is considerable personal variation in the effects of prominent word tone on voice quality. However, the effects of tone on voice quality are relatively consistent from syllables to words, although speaker S does change directions for her non-significant trends for the parameter OQ.

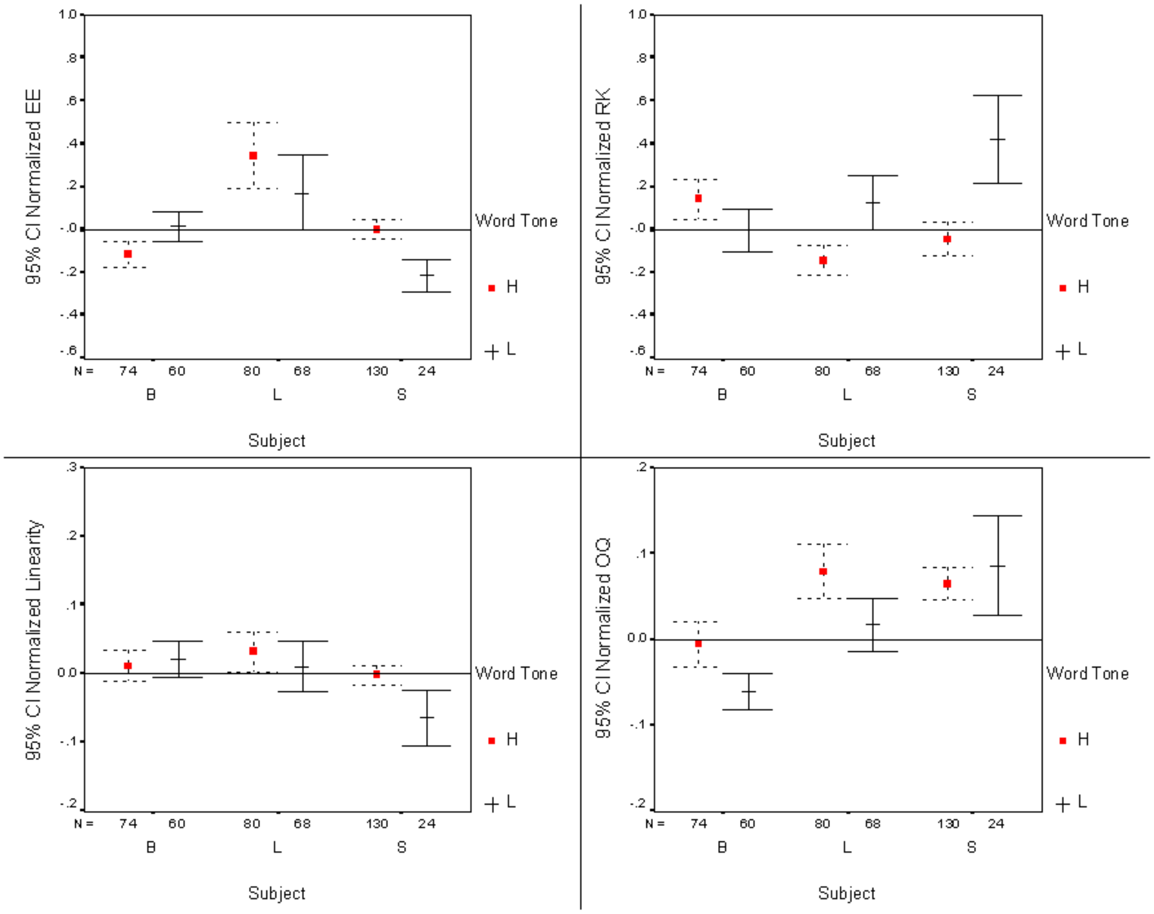


Figure 54. Means and 95% confidence interval error bars for tone of prominent words for the four Normalized LF measurements: EE (spectral intensity), RK (glottal symmetry), OQ (open quotient) and Linearity (spectral linearity).

Word Tone	B				L				S			
	EE	Lin	RK	OQ	EE	Lin	RK	OQ	EE	Lin	RK	OQ
H Tenser							X		X	X	X	
L Tenser	X			X				X				

Table 34. Summary of significant effects (solid 'X') of tone on the voice quality of prominent words.

5.4.3 Boundary tone of final syllable

A comparison of three samples (syllable-initial, -medial and -final) from sentence-final syllables did not find a consistent effect of boundary tone type across parameters or across speakers. The effects for all three speakers combined are summarized in Table 35. There were significant univariate effects for EE, OQ and Lin. However, two of these parameters, EE and RK, indicated that High boundary tones are tenser than Low boundary tones, and Linearity indicated the opposite. The MANOVA for the three speakers together did a significant main effect [F(1,217) = 23.982, $p \leq 0.001$]. There were, however, significant between-subject interactions for all four parameters: EE [F(2,213) = 68.626, $p \leq 0.001$], RK [F(2,213) = 27.045, $p \leq 0.001$], OQ [F(2,213) = 7.532, $p \leq 0.001$] and Linearity [F(2,213) = 30.988, $p \leq 0.001$].

Bound. Tone of Final Syllable		Normalized EE	Normalized RK	Normalized OQ	Normalized Linearity
H	Mean	.08887	.02833	.10118	.03488
	N	134	134	134	134
	Std. Deviation	.55402	.48847	.13226	.14297
L	Mean	-.18599	.08019	.05486	-.03797
	N	85	85	85	85
	Std. Deviation	.26408	.28712	.12408	.08675
	F(1,213)	34.192	.944	18.548	23.908
	p-value ≤	.001	.332	.001	.001

Table 35. Means, Ns, standard deviations, and *p*-values for boundary tones on final syllables for the four Normalized LF measurements, pooled across speakers: EE (spectral intensity), RK (glottal symmetry), OQ (open quotient) and Linearity (spectral linearity).

A review of the individual subjects' results, summarized in Figure 55 and Table 36 helps explain why there are inconsistent effects for the speakers as a group. One speaker, B, shows no effects of the boundary tone on the voice quality of the final syllable. Speakers L and S show opposite effects of boundary tone on the voice quality of the final

syllable. Furthermore, these effects are not consistent within the parameter RK (glottal skew). This parameter shows that High tones are tenser than Low tones for speaker L, but shows the opposite effect for speaker S. On the other hand, even though there are no consistent trends across speakers or within a parameter, the speakers themselves are consistent on what voice quality they use for boundary tones.

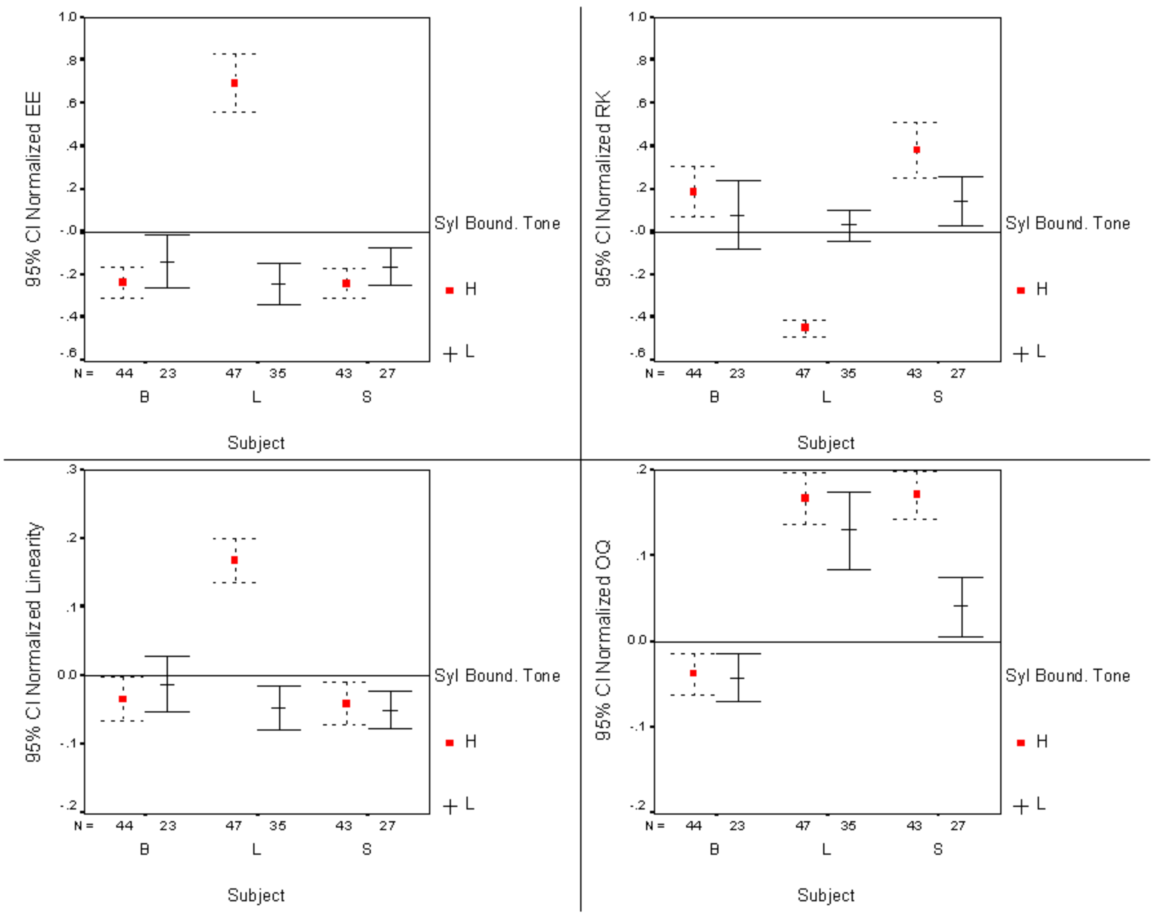


Figure 55. Means and 95% confidence interval error bars for boundary tone of final syllables for the four Normalized LF measurements: EE (spectral intensity), RK (glottal symmetry), OQ (open quotient) and Linearity (spectral linearity).

Boundary Tone	B				L				S			
	EE	Lin	RK	OQ	EE	Lin	RK	OQ	EE	Lin	RK	OQ
H Tenser					X	X	X					
L Tenser											X	X

Table 36. Summary of significant effects (solid 'X') of boundary tone on the voice quality of the final syllable.

5.4.4 Effects of phonological tone on non-modal phonation

Previous research has suggested that there is an association between F0 and the proportion of non-modal phonation or glottalizations occurring on a word, in that a low F0 is associated with an increase in the proportion of non-modal phonation (Hagen 1997). Figure 56 shows that in this corpus, speakers do have a greater tendency to produce non-modal phonation in sentences ending in Low boundary tones than in sentences ending with High boundary tones (chi-square = 45.761, $p \leq 0.001$). It is important to note, though, that speakers do also produce non-modal phonation on High boundary tones; Figure 57 shows that High boundary tones can have period doubling, noise and aperiodic waveforms.

On the other hand, as can be seen in Figure 58, speakers can produce non-modal phonation on both High prominent tone words and Low prominent tone words. Furthermore, a chi-square test for all the subjects combined shows that there is no significant difference in the proportion of non-modal phonation between High and Low prominent tones (chi-square = 1.603, $p \leq 0.205$). There is also not a consistent effect of prominent tone type on the proportion of non-modal phonation across the subjects either. Subject B produces non-modal phonation on both High and Low prominent tones, subject L produces non-modal phonation only on Low prominent tones, and subject S produces non-modal phonation only on High prominent tones.

In summary, non-modal phonation is associated with Low boundary tones. However, non-modal phonation is not associated with a general lowering of pitch, since it is not found more frequently on Low prominent tones than High prominent tones.

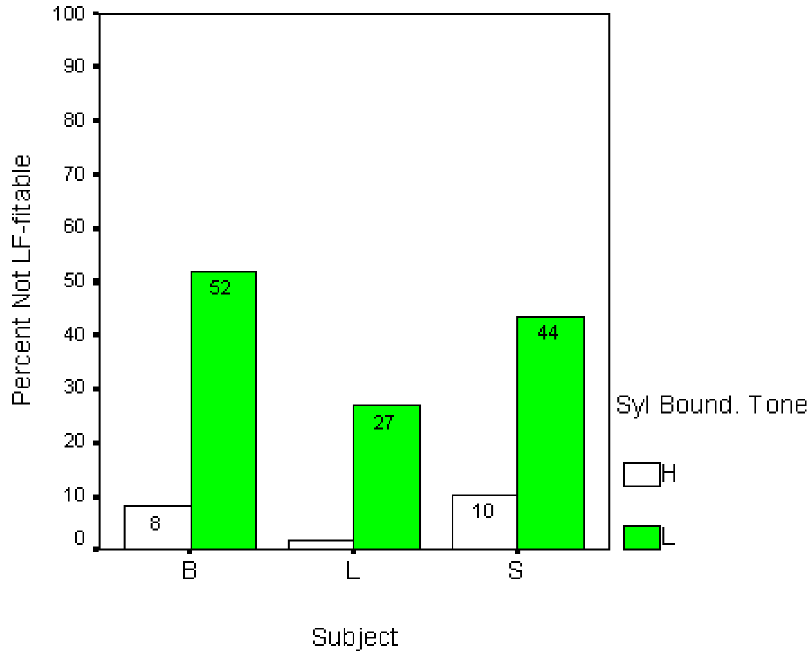


Figure 56. Percentage of non-LF-fitable waveforms for High tone and Low tone boundary syllables for each speaker.

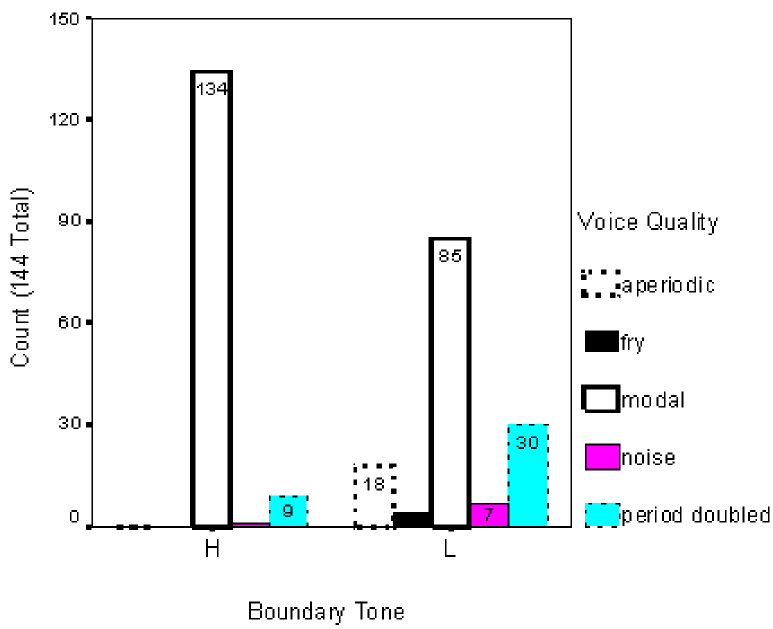


Figure 57. Percentage of different types of non-LF-fitable waveforms for High tone and Low tone boundary syllables.

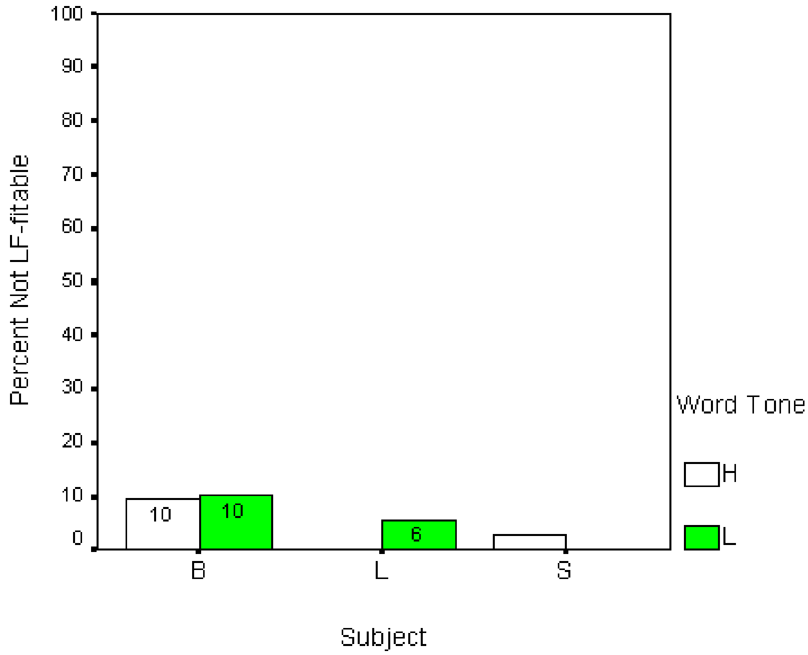


Figure 58. Percentage of non-LF-fitable waveforms for High tone and Low tone prominent words for each speaker.

5.4.5 Discussion of the effects of phonological tone on voice quality

The results presented above indicate that the effects of tone on modal voice quality differ greatly from speaker to speaker, and even within a single speaker, as can be seen in Table 37. Speakers can show significant differences for prominent tones, but not boundary tones (speaker B), or show opposite effects between prominent tones and boundary tones (speaker S), or even show significant effects in opposite directions on prominent tones (speaker L). Because of these trends, there is no clear consensus across the speakers as to whether High tones or Low tones have a tenser voice quality.

		Speaker B				Speaker L				Speaker S			
		EE	Lin	RK	OQ	EE	Lin	RK	OQ	EE	Lin	RK	OQ
Prominent Tone	H Tenser					X	(X)	X		X	X	X	
	L Tenser	X			X				X				
Boundary Tone	H Tenser					X	X	X					
	L Tenser											X	X

Table 37. Parameters used to indicate greater tenseness for High tones and Low tones separated by syllable and boundary tones and speaker. The parameters are: EE (spectral intensity), RK (glottal symmetry), OQ (open quotient) and Linearity (spectral linearity).

In retrospect, though, this trend of a lack of across-speaker and within-parameter effects of phonological tone on voice quality makes sense. Speakers produce both phonological high and low tones for prominent accents and phrase boundaries. If phonological tones had reliably different voice qualities across speakers or parameters, the previously discussed effects of phrase edges and prominence on voice quality would have been wiped out by the effects of tone.

Conversely, there are consistent effects of tone on the proportions of non-modal phonation found in the corpus. Speakers do produce significantly more non-modal phonation on Low boundary tones, indicating, as others have found, that “creaky voice” is associated with the ends of declarative sentences.

Chapter 6: General Discussion and Conclusion

As can be seen from the results presented in Chapter 5, there is a strong trend across subjects and across parameters to use changes in modal voice quality to distinguish between prominent and non-prominent words and to distinguish between phrase-initial prominent words and phrase-final prominent words. Subjects are also using non-modal phonation to distinguish between Low boundary tones and High boundary tones.

An overview of all the results, summarized in the tables below, brings out one more result. The parameter OQ (open quotient, glottal closing/closed duration) is used exclusively by speakers, both as a group and as individuals, to distinguish prominent words from non-prominent words in interrogatives, as can be seen in Table 38. OQ is also used by the speakers, as a group and by two of the speakers as individuals (B and L), to distinguish between prominent Low tones and prominent High tones. This combination of results indicates that a small value of OQ is a marker for Low prominent tones. In other words, Low tones, and thus prominent words in interrogatives, are marked by shorter glottal closing durations than High tones and prominent words in declaratives. The third speaker, S, shows a similar, but non-significant trend (see Figure 53). Speaker S also has the greatest proportion of High prominent tones in the interrogative sentences, at 72%. This suggests that speaker S's significantly different distribution of High and Low prominent tones may also be reflected in her personal voice quality tendencies for High and Low prominent tones and for prominent tones in declaratives and interrogatives.

There are also a number of other personal variations that occur in this study. For example, overall, speaker B makes the fewest distinctions with tense voice quality (as can be seen in the tables below). And, as noted in section 4.7 and 5.3.5, speaker B also produced the greatest proportion of non-modal phonation in the test and base corpora. As a result of this tendency, she may not need to make as many distinctions within the range of modal voice quality. However, it is important to note that producing a large proportion of non-modal phonation does not prevent speakers from using modal voice quality to make distinctions. For example, as can be seen in Figure 56, both speaker B and speaker S produce approximately equal amounts of non-modal phonation on the sentence-final syllable, *dads*. Speaker B, on the one hand, does not use modal voice quality to make a distinction between High and Low boundary tones. Speaker S, on the other hand, shows significant effects for boundary tones on modal voice quality – her Low boundary tones have a tenser voice quality than her High boundary tones. In short, the correlation between speaker B's high proportion of non-modal phonation and her tendency to make fewer distinctions using modal voice quality does not imply causation.

Another example of individual variation is in tendencies to show stronger effects for either declaratives or interrogatives for prominent words and phrase edges. Speaker L shows more significant effects for declaratives than interrogatives for both prominence and phrase edges. Speaker S, on the other hand, does not favor either declaratives or interrogatives. She does show strong effects of prominence primarily for interrogatives, but she shows equally strong effects for declaratives and interrogatives for phrase edges.

Furthermore, there is the possibility that these individual variations are reflections of speaker gender. As can be seen in Table 38, speakers B and S, the two females, do have similar trends for prominent words. Otherwise, individual variations are much more noticeable than variations along gender lines. Furthermore, there are only three speakers in this study, making it difficult to draw any conclusions, even if there had been stronger trends along gender lines.

Prominence	Speaker B				Speaker L				Speaker S			
	EE	Lin	RK	OQ	EE	Lin	RK	OQ	EE	Lin	RK	OQ
All Sentence Types	X	X		X	X				X		X	
Declaratives					X	X	X		X (dashed)			
Interrogatives	X	X	X	X		X (dashed)		X	X	X (dashed)	X	X (dashed)

Table 38. Parameters used to indicate greater tenseness for prominent words separated by sentence type and speaker. Significant effects are represented by a solid 'X' and near significant effects by a dashed 'X'. The parameters are: EE (spectral intensity), RK (glottal symmetry), OQ (open quotient) and Linearity (spectral linearity).

Phrase Edges	Speaker B				Speaker L				Speaker S			
	EE	Lin	RK	OQ	EE	Lin	RK	OQ	EE	Lin	RK	OQ
All Sentence Types			X (dashed)	X	X	X (dashed)	X	X	X	X	X	
Declarative-initial tenser				X	X	X	X	X (dashed)	X	X	X	
Interrogative-initial tenser								X (dashed)	X	X	X	

Table 39. Parameters used to indicate greater tenseness for phrase edges separated by sentence type and speaker. Significant effects are represented by a solid 'X' and near significant effects by a dashed 'X'. The parameters are: EE (spectral intensity), RK (glottal symmetry), OQ (open quotient) and Linearity (spectral linearity).

Phonological Tones		Speaker B				Speaker L				Speaker S			
		EE	Lin	RK	OQ	EE	Lin	RK	OQ	EE	Lin	RK	OQ
Prominent Tone	H Tenser					X	X (dashed)	X		X	X	X	
	L Tenser	X			X				X				
Boundary Tone	H Tenser					X	X	X					
	L Tenser											X	X

Table 40. Parameters used to indicate greater tenseness for High and Low tones separated by prominent tones, boundary tones and speaker. Significant effects are represented by a solid 'X' and near significant effects by a dashed 'X'. The parameters are: EE (spectral intensity), RK (glottal symmetry), OQ (open quotient) and Linearity (spectral linearity).

The next question is, why do speakers, as a group, use a tense voice quality in prominent and phrase-initial positions? As noted in Chapter 2, prominent and phrase-initial positions are positions that also receive articulatory and acoustic strengthening. Strengthening can be viewed as an increase in articulator effort for *particular* articulators. In other words, when a segment is strengthened, not all the speech articulators have an increase in effort, just the ones involved in that particular articulation. So for lingual consonants, for example, the tongue is raised higher and there is more contact on the palate (Fougeron 1999 citing Straka 1963). What would strengthening, however, mean for vowels? Several previous studies have found that there is an increase in allophonic vowel-initial word glottalization in strengthening (e.g., Pierrehumbert & Talkin 1992, Dilley et al. 1996, etc.), implying that vowels are become more consonantal in these positions. But for vowels in non-word-initial positions, contraction of several muscles within the larynx – the vocalis muscles and the interarytenoid muscles – will adduct the vocal fold and create a tense voice quality.

Moreover, de Jong (1995) hypothesized that an increase in strength of segments aids in lexical access. Vowels with tenser voice qualities acoustically have greater amplitudes of all harmonics in their source spectra, and the high frequency harmonics in particular. This can cause greater amplification of the formant frequencies. A pilot study using these subjects found that the amplitude of the high frequency harmonics is extremely low. They also have very fronted /u/s, making those F2s very high frequency. Thus, their second formant for /u/ in *doodads* is very low in amplitude. If they increase the amplitude of the high frequency harmonics in their source spectra, it could potentially aid listeners in lexical access of words containing front and fronted vowels.

Furthermore, speakers using a tense voice quality on prominent words can also assist listeners in discriminating phonological tone differences on prominent words. As Silverman's (in press) study of Jalapa Mazatec found, it is more difficult for listeners to discriminate phonetic pitch on breathy vowels than on modal voiced vowels. Also, cross-linguistically, languages do avoid making the full range of tonal contrasts on vowels with non-modal voice quality, e.g., languages tend not to contrast between breathy high and breathy low toned words (Silverman in press).[11] Accordingly, it is sensible for speakers to use a tense instead of lax voice quality on prominent words in English, because this will better enable listeners to perceive the phonological tonal contrasts that also appear on these words. It is not clear, though, from the results presented in Chapter 5, how this explanation would carry over to the discrimination of boundary tones on the final unaccented syllable of a sentence.

[11] Languages that do make these contrasts, such as Jalapa Mazatec, also do not maintain the non-modal voice quality throughout the entire word (Blankenship 1997; Silverman in press).

Finally, it is now possible to answer the first question in this study: Why is there glottalization on the word *cat* in the example sentence *I saw a cat.*? (See Figure 59.) The *cat* creaked because the sentence was a declarative and ended with a low boundary tone. The *cat* did *not* creak because *cat* was accented; speakers are more likely to produce non-modal phonation on non-prominent words. The *cat* did not creak solely because it was low-pitched; speakers are not more likely to produce non-modal phonation on low pitches in general.

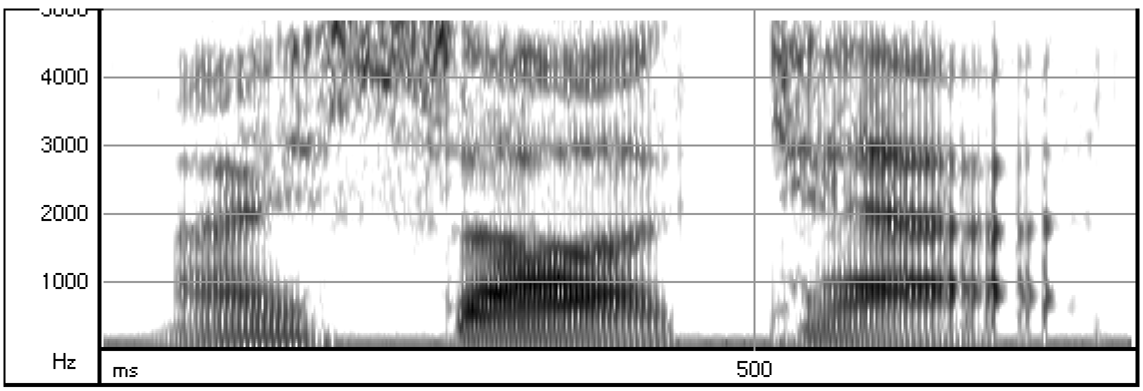


Figure 59. Spectrogram of the sentence *I saw a cat.* spoken with creaky voice on the final word.

In summary, if one is looking at the distribution of non-modal waveforms in a corpus, it is important to track the phonological tone of the boundary tones, but not necessarily the pitch accents. However, it is important to reiterate that there is a different distribution between allophonic vowel-initial word glottalizations and syllable-medial non-modal phonation. Vowel-initial glottalizations appear to be associated with tense voice quality and prosodic strengthening. Syllable-medial non-modal phonation is associated with phrase edges (i.e. sentence-initial *Da-* and sentence-final *dads*) and non-strengthened non-prominent words. This may indicate that there are physiological and/or perceptual

differences between the two types of *creak*, but the data collected from this and other acoustic studies cannot assess this hypothesis.

In conclusion, using the term “non-modal” for variations in voice quality can lead one to believe that the use of these variations is random and undesirable. Instead, it has been found that changes in voice quality occur in predictable positions – at phrase boundaries and on accented words. But there is still room for personal variations – particularly in the use of different voice qualities on the phonological tones in English. In general, though, when analyzing variations in modal voice quality predictable changes in voice quality can be found.

References

Baken, R. J. and R. F. Orlikoff. 2000. *Clinical Measurement of Speech and Voice*. Singular: San Diego.

Beckman, M. E. and G. M. Ayers. 1997. Guidelines for ToBI labelling, version 3. Ms. The Ohio State University.

Bickley, C. 1982. Acoustic analysis and perception of breathy vowels. *MIT Working Papers in Speech Communication* 1, 73-83.

Blankenship, B. 1997. *The Time Course of Breathiness and Laryngealization in Vowels*. Ph.D. dissertation, UCLA.

Byrd, D. 1992. Sex, dialects and reduction. *UCLA Working Papers in Phonetics* 81, 26-33.

Campbell, W. N. 1995. Loudness, spectral tilt, and perceived prominence in dialogues. *Proceedings of the XIIIth International Congress of Phonetic Sciences, Stockholm* 3, 676-679.

Childers, D. G. and C. Ahn. 1995. Modeling the glottal volume-velocity waveform for three voice types. *Journal of the Acoustical Society of America* 97(1), 505-519.

Cho, T. 2001. Effects of Prosody on Articulation. Ph.D. dissertation, UCLA. *UCLA Dissertations in Linguistics* 22.

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

de Jong, K. J. 1995. The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *Journal of the Acoustical Society of America* 97(1), 491-502.

Dilley, L. C. and S. Shattuck-Hufnagel. 1995. Variability in glottalization of word onset vowels in American English. *Proceedings of the XIIIth International Congress of Phonetic Sciences, Stockholm* 4, 586-589.

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

Epstein, M. A. 1999. *A Comparison of Linguistic and Pathological Breathiness Using the LF Model*. M.A. thesis, UCLA.

Epstein, M. A. 2001. Measuring voice quality in dynamic speech. *Poster presented at the 141st Meeting of the Acoustical Society of America, June 4-8, Chicago, Illinois*.

Epstein, M. A. and B. G. Payri. 2001. The effects of vowel quality and pitch on spectral and glottal flow measurements of the voice source. *Poster presented at the 141st Meeting of the Acoustical Society of America, June 4-8, Chicago, Illinois*.

Fant, G. 1960. *Acoustic Theory of Speech Production*. Mouton: The Hague, The Netherlands.

Fant, G. 1995. The LF-model revisited. Transformations and frequency domain analysis. *STL-QPSR* 2-3, 119-156.

Fant, G., J. Liljencrants and Q. Lin. 1985. A four-parameter model of glottal flow. *Paper presented at the French-Swedish Symposium, Grenoble, France*.

Fougeron, C. 1999. Prosodically conditioned articulatory variation: A review. *UCLA Working Papers in Phonetics* 97, 1-74.

Fougeron, C. and P. A. Keating. 1997. Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America* 101(6), 3728-3740.

Gerratt, B. R. and J. Kreiman. 2001. Toward a taxonomy of nonmodal phonation. *Journal of Phonetics* 29, 365-381.

Gobl, C. 1988. Voice source dynamics in connected speech. *STL-QPSR* 1, 123-159.

Gobl, C. and A. Ní Chasaide. 1988. The effects of adjacent voiced/voiceless consonants on the vowel voice source: A cross language study. *STL-QPSR* 2-3, 23-59.

Gobl, C. and A. Ní Chasaide. 1992. Acoustic characteristics of voice quality. *Speech Communication* 11, 481-490.

Gobl, C. and A. Ní Chasaide. 1999. Perceptual correlates of source parameters in breathy voice. *Proccedings of the XIVth International Congress of Phonetic Sciences, San Francisco*, 2437-2440.

Hagen, A. 1997. *Linguistic Functions of Glottalizations and their Language Specific Use in English and German*. Ph.D. dissertation, Friedrich-Alexander-Universität Erlangen-Nürnberg and MIT.

Hanson, H. M. 1997. Glottal characteristics of female speakers: Acoustic correlates. *Journal of the Acoustical Society of America* 101(1), 466-481.

Hanson, H. M. and E. S. Chuang. 1999. Glottal characteristics of male speakers: Acoustic correlates and comparison with female data. *Journal of the Acoustical Society of America* 106(2), 1064-1077.

Henton, C. G. and R. A. W. Bladon. 1985. Breathiness in normal female speech: Inefficiency versus desirability. *Language & Communication* 5(3), 221-227.

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

Herman, R. 1998. *Intonation and Discourse Structure in English: Phonological and Phonetic Markers of Local and Global Discourse Structure*. Ph.D. dissertation, The Ohio State University.

Herman, R., M. Beckman and K. Honda. 1996. Subglottal pressure and final lowering in English. *Proceedings of ICSLP 96* 1, 145-148.

Hirano, M. 1981. *Clinical Examination of Voice*. Springer-Verlag: Wien, Austria.

Holmberg, E. B., R. E. Hillman and J. S. Perkell. 1989. Glottal airflow and transglottal air pressure measurements for male and female speakers in low, normal, and high pitch. *Journal of Voice* 3(4), 294-305.

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

Huffman, M. K. 1987. Measures of phonation type in Hmong. *Journal of the Acoustical Society of America* 81(2), 495-504.

Huffman, M. K. 1999. Segmental and prosodic effects on coda glottalization. Ms. State University of New York, Stony Brook.

IEEE. 1969. IEEE recommended practice for speech quality measurements. *IEEE Transactions on Audio Electroacoustics* 17, 225-246.

Jakobson, R. and M. Halle. 1963. Tenseness and Laxness. In R. Jakobson, C. G. M. Fant, and M. Halle, eds., *Preliminaries to Speech Analysis: The Distinctive Features and their Correlates*. MIT Press: Cambridge, MA. 57-61.

Javkin, H. R. and I. Maddieson. 1983. An inverse filtering study of Burmese creaky voice. *UCLA Working Papers in Phonetics* 57, 115-125.

Keating, P. 1980. Patterns of fundamental frequency and vocal registers. In T. Murry and J. Murry, eds., *Infant Communication: Cry and Early Speech*. College-Hill Press: Houston, Texas. 209-233.

Kirk, P., J. Ladefoged and P. Ladefoged. 1984. Quantifying acoustic properties of modal, breathy, and creaky vowels in Jalapa Mazatec. *UCLA Working Papers in Phonetics* 59, 102-113.

Klatt, D. H. and L. C. Klatt. 1990. Analysis, synthesis, and pereception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America* 87(2), 820-857.

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

Kreiman, J. and B. Gerratt. 1996. The perceptual structure of pathologic voice quality. *Journal of the Acoustical Society of America* 100(3), 1787-1795.

Kreiman, J. and B. R. Gerratt. 2000. Sources of listener disagreement in voice quality assessment. *Journal of the Acoustical Society of America* 108(4), 1867-1876.

Kreiman, J. and B. Gerratt. 1998. Validity of rating scale measures of voice quality. *Journal of the Acoustical Society of America* 104(2), 1-11.

Kreiman, J., B. R. Gerratt, B. Gabelman, N. Antoñanzas-Barroso and M. Epstein. 2000. User's Guide: Analysis and Synthesis of Pathological Voice Quality. Unpublished technical report. UCLA.

Ladefoged, P. 1971. *Preliminaries to Linguistic Phonetics*. University of Chicago Press: Chicago.

Ladefoged, P. 1973. The features of the larynx. *Journal of Phonetics* 1(1), 73-83.

Ladefoged, P. 1983. The linguistic use of different phonation types. In D. Bless and J. Abbs, eds., *Vocal Fold Physiology: Contemporary Research and Clinical Issues*. College-Hill Press: San Diego, CA. 351-360.

Ladefoged, P. 1988. Discussion of phonetics: A note on some terms for phonation types. In O. Fujimura, eds., *Vocal Physiology: Voice Production, Mechanisms and Functions*. Raven Press: New York. 373-375.

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

Ladefoged, P., I. Maddieson and M. Jackson. 1988. Investigating phonation types in different languages. In O. Fujimura, eds., *Vocal Physiology: Voice Production, Mechanisms and Functions*. Raven Press, Ltd.: New York. 297-317.

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

Lehiste, I. 1975. The phonetic structure of paragraphs. In A. Cohen and S. G. Nooteboom, eds., *Structure and Process in Speech Perception*. Springer-Verlag: Heidelberg and New York. 195-203.

Lehiste, I. 1970. *Suprasegmentals*. MIT Press: Cambridge, Massachusetts.

Lieberman, P. 1977. *Speech Physiology and Acoustic Phonetics*. Macmillan: New York.

Löfqvist, A. and R. S. McGowan. 1992. Influence of consonantal environment on voice source aerodynamics. *Journal of Phonetics* 20, 93-110.

Maddieson, I. and S. Hess. 1987. The effect on F0 of the linguistic use of phonation type. *UCLA Working Papers in Phonetics* 67, 112-118.

Ní Chasaide, A. and C. Gobl. 1997. Voice Source Variation. In W. J. Hardcastle and J. Laver, eds., *The Handbook of Phonetic Sciences*. Blackwell: Oxford. 427-461.

Ohala, J. J. 1973. The physiology of tone. *Southern California Occasional Papers in Linguistics* 1, 1-14.

Pierrehumbert, J. 1995. Prosodic effects on glottal allophones. In O. Fujimura and M. Hirano, eds., *Vocal Fold Physiology: Voice Quality and Control*. Singular Publishing Group: San Diego. 39-60.

Pierrehumbert, J. B. 2000. #, +, and ? *Talk given at the Seventh Meeting on Laboratory Phonology, Nijmegen, Netherlands.*

Pierrehumbert, J. B. and S. Frisch. 1996. Synthesizing allophonic glottalization. In J. P. H. van Santen, R. W. Sproat, J. P. Olive, and J. Hirschberg, eds., *Progress in Speech Synthesis*. Springer: New York. 9-26.

Pierrehumbert, J. and J. Hirschberg. 1990. The meaning of intonation contours in the interpretation of discourse. In P. R. Cohen, J. Morgan, and M. E. Pollack, eds., *Intentions in Communication*. MIT Press: Cambridge, MA. 271-311.

Pierrehumbert, J. and D. Talkin. 1992. Lenition of /h/ and glottal stop. In G. J. Docherty and D. R. Ladd, eds., *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*. Cambridge University Press: Cambridge. 90-117.

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

Rothenberg, M. 1971. The glottal volume velocity waveform during loose and tight glottal adjustments. *Proceedings of the VIIth International Congress of Phonetic Sciences* , 380-388.

Schneiderman, C. R. 1984. *Basic Anatomy and Physiology in Speech and Hearing*. College-Hill Press: San Diego, CA.

Silverman, D. In press. Pitch discrimination during breathy versus modal phonation. In J. Local, R. Ogden, and R. Temple, eds., *Papers in Laboratory Phonology VI*. Cambridge University Press: Cambridge.

Slifka, J. 2000. *Respiratory Constraints on Speech Production at Prosodic Boundaries*. Ph.D. dissertation, MIT.

Sproat, R., M. Ostendorf and A. Hunt. 1999. The need for increased speech synthesis research. *Report of the 1998 NSF Workshop for Discussing Research Priorities and Evaluation Strategies in Speech Synthesis*.

Stevens, K. N. and H. M. Hanson. 1995. Classification of glottal vibration from acoustic measurements. In O. Fujimura and M. Hirano, eds., *Voice Quality Control*. Singular Publishing Group, Inc.: San Diego, CA. 147-170.

Straka, G. 1963. La divison de sons du language en voyelles et consonnes peut-elle être justifée? *Travaux de Ling.et de Littérature, U.de Strasbourg* 1, 17-99.

Titze, I. R. 1994. *Principles of Voice Production*. Prentice-Hall: Englewood Cliffs, NJ.

The UCLA Phonetics Laboratory. 1990. Dissection of the speech production mechanism. *UCLA Working Papers in Phonetics* 77. 1990.

Umeda, N. 1978. Occurence of glottal stops in fluent speech. *Journal of the Acoustical Society of America* 64(1), 88-94.

Vilkman, E., O. Aaltonen, U. Laine and I. Raimo. 1991. Intrinsic pitch of vowels – a complicated problem with an obvious solution? In J. Gauffin and B. Hammarberg, eds., *Vocal Fold Physiology: Acoustic, Perceptual, and Physiological Aspects of Voice Mechanisms*. Singular: San Diego, CA. 159-166.

Wightman, C. W., S. Shattuck-Hufnagel, M. Ostendorf and P. J. Price. 1992. Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America* 91(3), 1707-1717.