The Phonetics and Phonology of Glottal Manner Features

Mark Pennington

Submitted to the faculty of the University Graduate School
in partial fulfillment of the requirements
for the degree
Doctor of Philosophy
in the Department of Linguistics,
Indiana University

December 2005
Accepted by the Graduate Faculty, Indiana University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

**Doctoral Committee**

Robert F. Port, Ph.D., Linguistics, Chairman

Stuart Davis, Ph.D., Linguistics.

Kenneth de Jong, Ph.D., Linguistics

Charles S. Watson, Ph.D., Speech and Hearing Sciences

Gary Kidd, Ph.D., Speech and Hearing Sciences

**Date of Oral Examination**

November 18, 2005.
ABSTRACT

The purpose of this dissertation is threefold: i) to determine the number and acoustic-motor characteristics of the different phonation types, ii) to develop auditorily-based speech processing methods suitable for the measurement of glottal parameters, iii) to provide two equipollent pairs of glottal manner features that categorize the phonation types into the appropriate natural classes.

Nine pitch-independent phonation types appear necessary to account for linguistically significant contrasts: 1) glottal stop, 2) whisper, 3) breath, 4) harsh voice, 5) harsh whispery voice, 6) breathy voice, 7) tense voice, 8) plain voice, 9) lax voice. The phonation types (1, 2, 3) form the category of *glottal noise* while (2, 3) constitute *voicelessness*. The phonation types (4, 5, 6) are categorized as *noisy voice*, the phonation types (7, 8, 9) as *pure voice*. The three *glottal noise* phonation types (1, 2, 3) are characterized by aperiodic waveforms and exhibit increasingly larger glottal openings from 1) glottal stop to 3) breath. Because acoustic damping grows with widening glottal aperture, the first formant bandwidth (B1) likewise broadens. The three *noisy voice* phonation types (4, 5, 6) are characterized by periodic waveforms and have either considerable modulation noise (harsh voice), aspiration noise (breathy voice), or both (harsh whispery voice). The three *pure voice* phonation types (7, 8, 9) are also characterized by periodic waveforms, but with no significant modulation or aspiration noise, tense voice being cued by a flat spectral tilt, plain voice by an intermediate spectral tilt, lax voice by a steep spectral tilt.

The nine phonation types are classified both by I. a three-by-three motor hierarchy and II. a linear acoustic scale of derived glottal bandwidth (GBW) that
progressively narrows from 1) glottal stop to 9) lax voice. The primary motor features of position consist of the antagonistic pair [voice, noise]. The secondary motor features of stricture consist of the antagonistic pair [constricted, spread]. To demonstrate the generality of the feature framework adopted for glottal manner, equipollent feature systems for vowel height, backness, and supraglottal manner (major class features) are also proposed.
ACKNOWLEDGMENTS

I would like to thank all the members of my doctoral committee for their advice and assistance during the preparation of this dissertation. In particular, I am indebted to my advisor, Robert Port, not only for his insistence on paying attention to phonetic detail, but also for the enthusiastic discussions on the role of phonetics in phonology we have shared over the years. Stuart Davis has provided helpful feedback and suggestions throughout the course of this thesis while Charles Watson has been a source of inspiration for me on how to pursue scientific research.

James D. Miller’s questions on the linear acoustic scale of glottal bandwidth during my defense led me to consider additional ways of validating the concept, thereby benefiting this dissertation.

This research was funded in part by teaching assistantships in the Indiana University Linguistics Department.

I would like to express gratitude to my mother for her constant support and encouragement.
DEDICATION

In memory of Simone Elise
TABLE OF CONTENTS

Chapters:

1 Introduction ................................................................. 1

2 A survey of phonation types ............................................. 9
   2.1 Introduction .............................................................. 9
   2.2 Catford (1964, 1977) and Laver (1980) ......................... 10
      2.2.1 Catford (1964, 1977) ............................................. 10
      2.2.2 Laver (1980) .......................................................... 12
   2.3 Ladefoged (1971, 1973) and his co-workers (1996, 2001) ... 17
      2.3.1 Ladefoged (1971, 1973) ......................................... 17
      2.3.2 Ladefoged and co-workers (1996, 2001) ................. 22
   2.4 The inventory and organization of the phonation types ... 25
      2.4.1 The nine phonation types ..................................... 25
      2.4.2 The organization of the phonation types ................. 32
         2.4.2.1 Phonological evidence for a category of glottal noise ... 32
         2.4.2.2 Typological distribution of the phonation types .... 33
         2.4.2.3 Motor evidence for a hierarchy of phonation type .... 39
         2.4.2.4 Evidence for a linear acoustic scale of phonation type ... 47

3 Acoustic parameters of phonation type .......................... 52
   3.1 Introduction ............................................................ 52
   3.2 Laryngeal noise ........................................................ 53
      3.2.1 Mechanisms of modulation noise .......................... 53
      3.2.2 Measures of modulation noise .................................. 64
      3.2.3 Mechanisms of aspiration noise ............................ 68
      3.2.4 Measures of aspiration noise .................................. 76
   3.3 Spectral tilt .............................................................. 81
      3.3.1 Introduction ......................................................... 81
      3.3.2 Narrow-band spectral tilt: H1–H2 ......................... 82
      3.3.3 Mid-band and wide-band spectral tilt ..................... 84
   3.4 Summary and conclusion ........................................... 88
   Notes to Chapter 3 .......................................................... 91

4 Speech analysis methods ................................................. 92
   4.1 Introduction ............................................................. 92
   4.2 Pitch and voicing determination ................................. 93
      4.2.1 Auditory filter bank ............................................. 93
      4.2.2 F0 detection and the peak autocorrelation coefficient (pACC) ... 96
      4.2.3 Speech and voice detection .................................... 99
   4.3 Narrow-band slope (H1–H2) and estimation of modulation noise ... 100
   4.4 Wide-band harmonic slope and estimation of aspiration noise ... 101
      4.4.1 Introduction to Analysis-by-Synthesis Harmonic Estimation .... 101
4.4.2 Synthesis of model harmonic spectra................................. 102
4.4.3 The HNR and the slope of the short-time speech spectrum........103
4.5 First formant Q and the peak energy factor PE1............................. 104
4.6 Phonetically homogeneous subsegments.................................... 107
4.7 Summary.................................................................................. 109
Notes to Chapter 4........................................................................ 111

5 Evaluation of the speech analysis methods................................114
5.1 Assessment of pitch and voicing decisions.............................. 114
  5.1.1 Performance at very low fundamental frequencies.............. 114
  5.1.2 Pitch and voicing errors..................................................... 116
5.2 The log pACC and synthetic modulation noise......................... 118
5.3 Simultaneous jitter and shimmer............................................. 128
5.4 The HNR and synthetic aspiration noise................................. 133
  5.4.1 Evaluation of the ASHE algorithm using the model signals.... 133
  5.4.2 Evaluation of the ASHE algorithm using the glottal wave.... 142
5.5 Performance of the spectral tilt measures............................... 146
5.6 Comparison of the peak energy factor PE and the LP bandwidth.. 149
Notes to Chapter 5........................................................................ 156

6 Phonetic analysis of the phonation types................................. 159
6.1 Glottal noise: glottal stop, whisper, breath.............................. 159
  6.1.1 Standard Thai................................................................. 159
  6.1.2 Avar................................................................................. 161
  6.1.3 Quechua.......................................................................... 165
  6.1.4 Summary of the results for glottal noise............................ 175
6.2 Pure voice: tense voice and plain voice in Mpi........................ 177
6.3 Noisy voice: harsh voice, harsh whispery voice, breathy voice..... 180
  6.3.1 Gujarati............................................................................ 180
  6.3.2 Jalapa Mazatec................................................................. 187
6.4 General summary of the phonetic analysis of the phonation types.. 195
Notes to Chapter 6........................................................................ 200

7 Glottal manner features............................................................ 201
7.1 An overview of glottal manner feature systems....................... 201
  7.1.1 Introduction....................................................................... 201
  7.1.2 Trubetzkoy (1938) and the complex segment..................... 201
  7.1.3 Jakobson, Fant, and Halle (1952)....................................... 209
  7.1.4 Chomsky and Halle (1968)............................................... 211
  7.1.5 Halle and Stevens (1971).................................................... 214
  7.1.6 Recent systems of glottal manner features....................... 226
7.2 The present system of glottal manner features....................... 235
  7.2.1 The acoustic-motor classification of the nine phonation types.. 235
  7.2.2 The proposed system of glottal manner features.................. 237
    7.2.2.1 Equipollent motor features and voice assimilation......... 237
    7.2.2.2 Natural classes of phonation types.............................. 243
7.2.2.3 Equipollent perceptual and gestural markedness………247

7.3 The feature framework and other speech parameters..........................253
  7.3.1 Height features.............................................................................253
  7.3.2 Backness features.........................................................................256
  7.3.3 Supraglottal manner features (major class features)..................257
  7.3.4 Organizing principles of the four feature subsystems...............262

7.4 General summary and directions for future research.........................264

Notes to Chapter 7..................................................................................268

Appendix A: analysis of the Mpi speech samples......................................273
Appendix B: analysis of the Gujarati speech samples.............................281
Appendix C: analysis of the Jalapa Mazatec speech samples.....................296

References..............................................................................................305
CHAPTER 1
INTRODUCTION

The goal of this dissertation is to examine the motor and acoustic properties of the different phonation types and then to determine a system of phonological features that best corresponds to their phonetic characteristics. A good deal of effort has been devoted to developing and testing speech analysis methods to quantify the acoustic manifestations of phonation type. The main impetus for this bottom-up approach is the belief that the discovery of the appropriate acoustic parameters may help to constrain hypotheses about speech mechanisms and ultimately guide the choice of phonological representations.

The beginning of Chapter 2 introduces the fundamental notions of phonation type by reviewing the inventories most often used in the world’s languages. The work of Catford (1964, 1977) constitutes the basis of many of the later studies on phonation type. He offers motoric and acoustic definitions for breath, whisper, voice, glottal stop, breathy voice, creaky voice, whispery creak and mentions a number of languages in which the phonation types may be lexically contrastive. Laver (1980) expands upon his work by bringing further evidence to bear on the inventory of phonation types. His contribution includes elaborating the concept of compound phonation types like harsh whispery voice, in which the vocal folds are both constricted and spread, as well as cross-classifying the phonation types by pitch register. Taking a more linguistically-oriented approach, Ladefoged (1971, 1973) considers the different phonation types to be formed by a continuum of glottal stricture going from breath, the most open position of the vocal
folds, to *glottal stop*, the closed position, with voice situated centrally between the two extremes.

After the phonological distinction between *breath* and *whisper* is established on the basis of the onset consonants preceding Proto-Tai tones, nine pitch-independent phonation types are postulated, all of which are phonologically pertinent with the possible exception of *harsh whispery voice*:

1) glottal stop
2) whisper
3) breath
4) harsh voice
5) harsh whispery voice
6) breathy voice
7) tense voice
8) plain voice
9) lax voice

Cross-linguistic evidence is provided suggesting that the phonological category *glottal noise* includes the three phonation types *glottal stop*, *whisper*, and *breath*, while *whisper* and *breath* constitute *voicelessness*. Because the *glottal stop* is made with either a complete, or near closure of the glottis (e.g. during the attack and release phases), the damping caused by glottal losses is small. Additional damping is brought about by the wider glottal aperture of *whisper*, and even more so by the fully open glottis of *breath*. If other speech parameters remain constant, an increase in glottal damping leads to a reduction in the first formant quality factor (Q1).
The *noisy voice* phonation types all have a periodic glottal source mixed with a significant amount of aperiodicity: *harsh voice* with considerable period-to-period irregularity (modulation noise), *breathy voice* with considerable turbulence noise (aspiration noise), and *harsh whispery voice* with both kinds of noise at the same time. 

The *pure voice* phonation types are characterized by negligible modulation and aspiration noise, *tense voice* with a spectral tilt significantly flatter than plain voice, *plain voice* with a spectral tilt lying between lax voice and tense voice, and *lax voice* with a spectral tilt significantly steeper than plain voice.

The nine phonation types appear to be arranged motorically in a three-by-three fashion. The primary motor dimension of position is characterized by opposing action between an extreme vocal fold position (*glottal noise*) and a central one (*pure voice*), the central-extreme position (*noisy voice*) resulting from the simultaneous implementation of *glottal noise* and *pure voice*. The secondary motor dimension of stricture involves antagonistic action between a constricted vocal fold stricture and a spread one, the constricted-spread gesture being due to the simultaneous implementation of the constricted and spread vocal folds, e.g. during *whisper, harsh whispery voice*, and *plain voice*. In view of the typological evidence, the nine phonation types also appear to be ordered acoustically along a linear scale of derived glottal bandwidth (GBW) from 1) *glottal stop* with the largest GBW to 9) *lax voice* with the narrowest GBW.

Chapter 3 considers in detail the motor mechanisms underlying modulation and aspiration noise as well as the methods currently used to estimate their relative levels. Modulation noise arises from the cycle-to-cycle perturbation of the pitch period resulting from the greater than normal adduction of the nonlinear vocal fold system. Aspiration
noise is generated by distributed turbulent sources in the region of the larynx and is characterized by a relatively flat spectrum when the radiation at the lips is taken into account. Spectral tilt and its associated acoustic measures are also discussed in light of the properties of the glottal waveform.

Chapter 4 describes the auditorily-based speech processing algorithms used in this work. A bank of second-order linear filters roughly models the response of the auditory periphery. The autocorrelation of the sum of the low-frequency filter outputs yields estimates of F0 and the modulation noise (the log peak AutoCorrelation Coefficient or log pACC). The narrow-band spectral tilt (H1–H2) is calculated automatically on the basis of this F0 estimate. The aspiration noise (the harmonics-to-noise ratio or HNR) and the wide-band spectral tilt (the harmonic slope) are estimated simultaneously by means of a least squares Analysis-by-Synthesis Harmonic Estimation procedure (ASHE). A method of approximating first formant Q1, termed the peak energy factor PE1, is likewise developed. Lastly, a set of segmentation criteria is put forward to determine the boundaries of phonetically homogeneous subsegments in the speech wave.

Chapter 5 is dedicated to evaluating the speech analysis methods detailed in Chapter 4. The autocorrelation method used for pitch and voicing determination shows fewer voicing errors than a standard cross-correlation pitch detection algorithm. There is, however, a slightly greater incidence of gross pitch errors. To evaluate the log pACC as an acoustic measure of voice source irregularity, signals are constructed with varying amounts of synthetic modulation noise (jitter and shimmer). The results show that there are very high correlations between the log pACC and jitter (about −0.98) while the associations between the log pACC and shimmer are only slightly weaker (on the order
of –0.96). These strong linear relationships suggest that the log peak autocorrelation coefficient is the most natural and economical parameter for characterizing modulation noise since it is already used in the determination of F0 and voicing.

The harmonics-to-noise ratio (HNR) and the harmonic slope, as measured by the ASHE algorithm, are likewise evaluated with synthesized waveforms. Most often there is a nearly one-to-one association between the actual and mean HNRs. Furthermore, the HNR estimates of the ASHE technique are shown to be insensitive to fairly large amounts of modulation noise. As a result, aspiration noise is nearly always kept distinct from modulation noise, and vice versa, which is important when noisy voice phonation types are present in the speech signal, i.e. harsh voice, harsh whispery voice, and breathy voice. In addition, the measured harmonic slopes are very similar to the synthesized ones. Even at the lowest synthesized HNR of 0.75 dB, the standard deviations of the measured harmonic slopes show a relatively small degree of scatter around the mean.

The peak energy factor PE, as defined in Chapter 4, has three properties in common with the quality factor Q: it is dimensionless, invariant with respect to multiplicative frequency shifts, and serves as an indicator of the amplification at resonance. Linear prediction is the current speech processing method for determining formant bandwidths, and thus formant Qs. The PE is compared with the linear prediction method by using signals with known resonant frequencies and bandwidths. The measured values of both the PE and linear prediction bandwidth all display strong linear relationships with the synthetic ones. The peak energy factor PE may therefore be viewed as a more auditorily plausible and computationally simpler alternative to linear predicative bandwidth estimation methods.
In Chapter 6 several languages with known phonemic contrasts in phonation type are investigated. The speech analysis procedures developed in Chapter 4 appear to distinguish rather well among the nine phonation types, each of which is represented by several particularly good exemplars in the present language data. When postvocalic /ʔ/ and /h/ are compared pairwise in Avar, the PE1s are higher for /ʔ/ than they are for /h/, in conformity with the assumption that the F1 of closed glottis /ʔ/ is subject to less damping than the F1 of open glottis /h/. Quechua has a three-way contrast between voiceless, aspirated voiceless, and ejective obstruents. Because the ejective is generated by an initiatory glottal stop, its supraglottal release should be accompanied by a larger PE1 value than those of the voiceless and aspirated voiceless stops. As expected, the uvular ejective, the velar ejective, and the fricative portion of the palato-alveolar affricate ejective display higher PE1s than the corresponding voiceless and aspirated voiceless obstruents. The unaspirated velar and the fricative portion of the unaspirated palato-alveolar affricate likewise have higher PE1s than the aspirated ones, indicating a three-way contrast of increasing glottal opening at these two places of articulation, i.e. glottal stop, whisper and breath. Mpi illustrates the phonological opposition between plain voice and tense voice. Gujarati is characterized by a marginal phonemic contrast between pure and breathy voice, while Jalapa Mazatec has a three-way opposition between pure voice, harsh voice, and breathy voice.

Chapter 7 first provides critical reviews of earlier systems of glottal manner features, i.e. those of Trubetzkoy (1938), Jakobson et al. (1952), Chomsky and Halle (1968) as well as Halle and Stevens (1971). The more recent systems of Kenstowicz (1994) and Lombardi (1994) are then examined. The phonetic and phonological
difficulties that arise from the use of privative features are highlighted and several reasons are adduced as to why equipollent features are to be preferred. Taking Halle’s 1983 motor premise as a starting point, phonological features are assumed to be equipollent and implemented through the coordinated action of antagonistic muscle groups.

The nine phonation types are classified according to I. a three-by-three motor hierarchy and II. an acoustic ordering along a linear scale of derived glottal bandwidth (GBW) that progressively narrows from 1) *glottal stop* to 9) *lax voice*.

The glottal manner features of the primary motor dimension of position consist of the antagonistic pair: [noise] and [voice].

A) *Glottal noise* is indicated by the feature set [+noise, –voice]

B) *Noisy voice* by [+noise, +voice],

C) *Pure voice* by [–noise, +voice].

The glottal manner features of the secondary motor dimension of stricture consist of the antagonistic pair: [constricted] and [spread].

The contrasts within *glottal noise* are:

1) *Glottal stop* [+noise, –voice, +constricted, –spread],

2) *Whisper* [+noise, –voice, +constricted, +spread],

3) *Breath* [+noise, –voice, –constricted, +spread].

The contrasts within *noisy voice* are:

4) *Harsh voice* [+noise, +voice, +constricted, –spread],

5) *Harsh whispery voice* [+noise, +voice, +constricted, +spread],

6) *Breathy voice* [+noise, +voice, –constricted, +spread].
The contrasts within *pure voice* are:

7) *Tense voice* \([-\text{noise}, +\text{voice}, +\text{constricted}, -\text{spread}]\),

8) *Plain voice* \([-\text{noise}, +\text{voice}, +\text{constricted}, +\text{spread}]\),

9) *Lax voice* \([-\text{noise}, +\text{voice}, -\text{constricted}, +\text{spread}]\).

The above system allows for the characterization of a number of natural classes of glottal manner. The perceptual and gestural markedness of the simple three-term equipollent opposition is discussed with a view toward explaining why phonemic *harsh whispery voice* \([+\text{voice}, +\text{noise}, +\text{constricted}, +\text{spread}]\) seldom if ever occurs. Only the system of glottal manner features displays a middle term with two positively specified equipollent features, as in the combinations \([+\text{noise}, +\text{voice}]\) and \([+\text{constricted}, +\text{spread}]\). The other feature subsystems show evidence of middle terms with negatively specified equipollent features \([-\text{F}_1, -\text{F}_2]\). In order to demonstrate the generality of the feature framework adopted for glottal manner, equipollent feature systems for height, backness, and supraglottal manner (major class features) are also proposed.
CHAPTER 2
A SURVEY OF PHONATION TYPES

2.1 Introduction

The present chapter is devoted to a review of the various phonation types. The phonetic categories treated will be those mainly associated with phonological distinctions in a wide range of genetically diverse languages, as in the approach of Ladefoged and his colleagues. Nevertheless, Laver’s compound phonation types such as *harsh whispery voice*, which never appear to enter into lexically contrastive oppositions according to the available language descriptions, will be discussed in so far as these phonation types are indicators of sociolinguistic habits and convey paralinguistic meaning (for a review of the literature on the sociolinguistic and paralinguistic use of phonation types, see Esling, 1978; Henton and Bladon, 1988). Following Sweet’s remarks (1877: 98) on the harsh voice of Scotsmen and Saxon Germans, Laver (1980: 131-132) points out that some urban Scots accents are frequently characterized by ventricular or possibly whispery ventricular voice. He also notes that paralinguistic signals of anger are mediated by harsh or ventricular voice in English. As harsh voice being used to express anger represents no less a linguistic choice than when it is being used phonemically, the inventory of phonation types should reflect this fact and include it. A similar situation arises with regard to the use of distinct pitch contours (rising, falling, level) to instantiate either phonemic tone or paralinguistic intonation depending on the particular language. After a review of the work over the last forty years or so, our proposals for the inventory and the organization of the phonation types are put forward and discussed.
2.2 Catford (1964, 1977) and Laver (1980)

2.2.1 Catford (1964, 1977)

In his 1977 classification of phonation types, Catford proposed four degrees of glottal stricture and four combinations of these. He also recognized four different locations within the larynx: full glottal, anterior (called ligamental in 1964), posterior (called arytenoidal in 1964), and ventricular.

Taking into account the additional stop stricture type of his 1964 paper, the five degrees of glottal stricture are:

1) **Breath** glottis wide open, area 60-95% the maximum value, with a critical volume velocity of 200-300 cm³/s required to produce sound.

2) **Whisper** glottis constricted, area less than 25% the maximum value, a critical volume velocity of 25-30 cm³/s.

3) **Voice** periodic vibration of the vocal folds, fundamental frequency from 60-70 to 1100 Hz, typical mean glottal area on the order of 0.1 cm², a critical volume velocity of 50 cm³/s, typical volume velocity 100-350 cm³/s; *anterior* or *ligamental voice* (also called *sharp* or *tense voice*) is produced with the arytenoids clamped tightly together and may involve a sphincteric constriction of the upper larynx as well.

4) **Creak** periodic vibration of a small section of the vocal folds near the thyroid cartilage, very low fundamental frequency 40-50 Hz, low volume velocity 12-20 cm³/s.

5) **Glottal stop** glottis tightly closed, built up positive or negative pressure that can be released explosively into breath, whisper, voice, or creak.
The combined phonation types are:

6) *Breathy voice* combination of the turbulent airflow of breath and the vibration of voice, vocal folds do not meet in the center, they “flap in the breeze”, high volume velocity 900-1000 cm³/s.

7) *Whispery voice* combination of the narrowed glottis of whisper and the vibration of voice (also called *murmur*), the first type involves fully closing anterior vocal fold vibration with whisper escaping through a chink between the arytenoids, in the second type the relaxed vocal folds never close completely, with whisper escaping through a chink somewhere along their length, volume velocity from 60 to 300-400 cm³/s.

8) *Whispery creak* combination of whisper and creak.

9) *Creaky voice* combination of creak and voice (alternatively called *voiced creak* in 1964).

Whereas many languages have only a two-way contrast in phonation type breath ~voice, Catford (1964) observes that it is not unusual to find three-way phonological contrasts as well. For example, Danish has breath ~ voice ~ voiced creak, one variety of Hindi breath ~ voice ~ whispery voice, another variety of Hindi breath ~ ligamental voice ~ whispery voice, Nilotic breath ~ voice ~ ligamental voice. In some varieties of Gujarati there is a possible four-way contrast breath ~ voice ~ ligamental voice ~ whispery voice.
2.2.2 Laver (1980)

To determine whether there were any systematic constraints governing phonatory settings, Laver (1980: 111-118) examined their combinability and concluded that there are three broad categories:

1) The first category consists of *modal voice* and *falsetto*. Each of these voiced phonatory settings can occur alone as simple types and combine with the members of the other two categories, but they can not be present simultaneously.

2) The second category includes *creak* and *whisper*. Both can occur alone as simple types, and together as the compound type *whispery creak*. Each member of the second category can combine with each member of the first category to yield the compound types *whispery voice*, *whispery falsetto*, *creaky voice*, *creaky falsetto*. Not only can members of the second category combine with members of the first category, but they can also combine with each other at the same time, yielding the compound phonatory settings *whispery creaky voice*, *whispery creaky falsetto*.

3) The members of the third category, *harshness* and *breathiness*, can not occur alone, and can therefore only modify other phonatory types. Combined with members of the first category, the two modificatory settings yield the compound phonatory settings *harsh voice*, *harsh falsetto*, *breathy voice*. In addition, each can not modify the members of the second category alone, but only the members of the first and second categories taken together. Hence they form the compound phonatory settings *harsh whispery voice*, *harsh whispery falsetto*, *harsh creaky voice*, *harsh creaky falsetto*, *harsh whispery creaky voice*, *harsh whispery creaky falsetto*. 


In the course of his exposition, Laver did not stress the distinction between those phonatory settings resulting from an adjustment in fundamental frequency and those due to a particular type of glottal stricture. However, if the three categories are looked at in more detail, they do seem to be organized on the basis of a distinction between pitch register and phonation type. First, it is clear that Laver distinguishes three pitch registers:

1) *Falsetto* (high register)
2) *Modal* (mid register)
3) *Creaky* (low register)

Furthermore, he assumes two modificatory settings *harshness* and *breathiness*, the logical combinations of which yield four voiced phonation types independent of pitch register:

1) *Pure* (neither *harsh* nor *breathy*)
2) *Harsh*
3) *Breathy*
4) *Harsh whispery* (both *harsh* and *breathy*)

When the three pitch registers are combined orthogonally with the four voiced phonation types, the twelve possibilities represent his classification of the voiced phonatory settings fairly well.
Table 2.1 shows that 1) pitch registers do not occur simultaneously and 2) phonation types combine freely with each of the three pitch registers. In support of the first statement, note that the two exclusive members of Laver’s first category, *modal voice* and *falsetto* span nearly discontinuous pitch ranges. Citing a 1968 study by Hollien and Michel, Laver (1980: 118-119) noted that the average pitch range was 94-287 Hz for male modal voice, whereas it was 275-634 Hz for male falsetto. Furthermore, the motor characteristics of modal voice and falsetto are also quite different. Modal voice is produced with moderate adductive tension of the interarytenoid muscles, moderate medial compression of the ligamental glottis by the lateral cricoarytenoid and thyroarytenoid muscles, and moderate longitudinal tension of the vocalis and cricothyroid muscles. Falsetto is generated with high adductive tension, large medial compression, and high passive longitudinal tension of the vocal folds through contraction of the cricothyroid muscles. Since the vocal folds can not be stretched lengthwise to achieve higher pitch and yet simultaneously be shortened and thickened for lower pitch, it is easy to understand why the *modal voice* and *falsetto* pitch registers are subject to a co-occurrence restriction (1980: 137).

Because the difference of pitch between the *creaky* and *falsetto* registers is even more pronounced than the one existing between the *modal voice* and *falsetto* registers,
Laver’s compound phonatory setting _creaky falsetto_ should not exist. If, on the other hand, _creak_ is thought of as an impressionistic label for a kind of harshness, then _creaky falsetto_ would be one realization of _harsh falsetto_ (cf. also 1980: 138). The latter compound phonatory setting is in complete conformity with the second statement, as it consists of the _harsh_ phonation type freely combining with the _falsetto_ pitch register.

Both Laver and Catford make a distinction between _creak_ and _creaky voice_, though it is not clear from their respective descriptions what physical parameters keep them apart. Ladefoged (1971: 15) mentions that it is doubtful whether such a distinction is necessary for a theory of linguistic phonetics. Laver (1980: 122-126) points out that the mean F0 of _creak_ for male speakers is 34.6 Hz, with a range of 24 to 52 Hz. He cites an excerpt from a study on vocal fry by Hollien et al. (1966) where not only are the vocal folds are apparently compressed, but the ventricular folds are somewhat adducted as well, the lower surfaces of the ventricular folds coming into contact with the upper surfaces of the vocal folds and forming “an unusually thick, compact, (but not necessarily tense) structure.” Summarizing the work of Hollien, his co-workers, and other researchers, Laver states that _creak_ is produced by “strong adductive tension and medial compression, but little longitudinal tension, and with vigorous ventricular involvement.”

Similarly to _modal voice_, the mean F0 of _harsh voice_ is 122.1 Hz. However, unlike _modal voice_ and somewhat more similarly to _creak_, _harsh voice_ is characterized by extreme adductive tension and medial compression, with an equally extreme degree of ventricular participation, hence the frequent designation _ventricular voice_ for severely _harsh voice_. Laver describes the modificatory harshness setting as follows (1980: 118):

“…the ventricular folds become involved in the phonation of the true vocal fold by squeezing closed the ventricle of Morgagni and pressing down on the true
vocal folds, with the effect that the true and the ventricular folds combine to vibrate as more massive composite elements. In order to bring the ventricular folds to this position, a high degree of muscular tension is needed, and the effect is to make phonation auditorily very harsh.”

Additionally, Laver finds that the acoustic correlate of harshness is a highly irregular acoustic waveform displaying a great deal of jitter or period-to-period variability (see Michel, 1964 in particular).

Like Catford, Laver posits the combined phonation types breathy voice and whispery voice. Note, however, that Catford’s second type of whispery voice can grade imperceptibly into breathy voice; his first type may do so as well when the interarytenoid gap widens (cf. Laver, 1980: 135). Moreover, no language uses these two types in phonemically contrastive pairs, probably because of the lack of a robust perceptual difference between them. Therefore, whispery voice will be considered a narrowed glottis variant of phonological breathy voice. Phonetic whispery voice occurs in all phonatory settings requiring a relatively narrowed glottis for their realization, such as whispery creaky voice, whispery falsetto, and harsh whispery voice. Addressing the issue of the compatibility of the motor gestures that make up the phonatory settings, Laver (1980: 136) states:

“Whisper is a case in point, where low adductive tension is normally one of its characteristics: it does mean though that the whisper component in harsh whispery voice, say, is maintained by a much greater effort on the part of the...cricoarytenoid muscles to keep the arytenoid triangle open against the vigorous attempt by the interarytenoid muscles to close it.”

The conclusion of this section on Laver will focus on his view of tension. Recall that Catford’s inventory of phonation types included one form of voice produced with a tense constriction of the anterior or ligamentary vocal folds. Laver, on the other hand, claims that the tension parameter likewise involves the resonance properties of the
supraglottal vocal tract, thereby excluding it from the system of phonation types, strictly speaking. Nevertheless, Laver (1980: 142) recognizes the importance of the spectral tilt as an indicator of tension. For example, he sums up the early work of Van Dusen (1941): “…tense and lax voices seem to be acoustically differentiated chiefly by the relative amounts of energy in the upper harmonics, tense voice having stronger upper harmonics than lax voice.” He also mentions the sharp and soft voices of Chiba and Kajiyama (1958) defined in a very similar fashion. As will be discussed later, the bandwidths of the vocal tract resonances do become larger with increasing glottal opening. Yet because the vocal folds are narrowly approximated during voicing, the damping is small in comparison to open glottis breath and whisper. Bandwidth is therefore a weak cue for tension on voiced sounds.

2.3 Ladefoged (1971, 1973) and his co-workers (1996, 2001)

2.3.1 Ladefoged (1971, 1973)

To introduce the states of the glottis, Ladefoged (1971: 8) provides a list of those he considers linguistically pertinent together with a brief description:

1) Voiceless vocal folds apart.

2) Whisper vocal folds together or narrowed except between the arytenoids.

3) Murmur “breathy voice”, arytenoids apart, ligamental vocal cords vibrating.

4) Voice vibration of the vocal cords.

5) Laryngealisation “creaky voice”, arytenoids tightly together, but a small length of the ligamental vocal cords vibrating.

6) Glottal stop vocal cords held together.
7) Aspiration a brief period of voicelessness during and immediately after the release of the articulatory stricture.

Unlike the other phonation types, aspiration lies outside the glottal stricture ranking since it involves timing considerations as well.

Voiceless (breath in Catford’s terminology) is produced with the arytenoids pulled as far apart as during normal expiration. Turning to the distinctive function of voiceless, Ladefoged (1971: 11) notes the existence of phonemically contrastive voiceless stops and fricatives (very frequent), nasals and laterals (Burmese), and glides (Scottish English). He also states that voiceless vowels often occur as allophones of their voiced counterparts, but has serious doubts as to their phonemic use (see also Gordon, 1998 on this question).

Whisper occurs when “the vocal cords are narrowed or even together anteriorly, leaving a somewhat wider gap at the other end between the arytenoid cartilages. The cords are held rather stiffly, and sometimes there are additional constrictions just above the glottis” (1971: 16). Ladefoged mentions a possible contrast in French and Wolof between whispered and voiceless (i.e. breath) sounds in the environment of a pause, apparently when the whispered sound is an allophone of a voiced sound, but he does not offer further details (1971: 19).

Murmur (Catford’s whispery voice) takes place when “the posterior portions (between the arytenoid cartilages) of the vocal cords are held apart, while the ligamental parts are allowed to vibrate. There is a high rate of flow of air out of the lungs; so the term ‘breathy voice’ is also appropriate.” The term murmur was first used by Pandit (1957) to describe phonemically breathy Gujarati vowels. In most Indo-Aryan languages,
there is a four-way series of stops consisting of voiceless, voiced, voiceless aspirated, and
those with a murmured release (voiced aspirated). Phonemic murmured stops and nasals
are frequent in Southern Bantu, but are realized with less aspiration noise than their Indo-

The terms *laryngealized* and *laryngealization* were coined by Pike (1943: 127) to
refer to glottal trillization with superimposed voice. Ladefoged (1971: 14-15) defines a
laryngealized sound in much the same way as Catford does *creak*: vibration of the
compressed anterior (ligamental) vocal folds, yielding “often a harsh sound with
comparatively low pitch.” Nonetheless, in his description of Lango, a Nilotic language
contrasting laryngealized and nonlaryngealized vowels, he states that tones can occur
with either kind of phonation, thereby implying a certain independence of laryngealized
sounds relative to pitch level. In the Chadic languages such as Hausa, Bura, and Margi,
there is also an opposition between laryngealization and normal voicing on stops and
glides. As certain types of laryngealized and breathy voice seem to require an even finer
division of the glottal stricture continuum, Ladefoged (1971: 17) splits *murmur* into
*breathy voice, murmur, lax voice* and *laryngealization* into *tense voice, creaky voice,
creak*:

1) Voiceless
2) Breathy voice
3) Murmur
4) Lax voice
5) Voice
6) Tense voice
7) *Creaky voice*

8) *Creak*

9) *Glottal stop*

With no language making more than three phonological contrasts on this continuum, and with most languages making only two, he (1971: 18) hypothesizes that:

“...there is a continuum extending from the most closed position, a glottal stop, to the most open position observed in speech, which is that in voiceless sounds. Starting from a *glottal stop* (which itself may have several degrees of tightness), it is possible to pass through a form of laryngealization (here called *creak*) in which the arytenoid cartilages are pulled toward one another, and the whole glottis remains constricted except for a small opening in the anterior portion. Slight (but not complete) relaxation of the pulling together of the arytenoids produces the next phonation type, *creaky voice*, in which a larger proportion of the glottis is vibrating. Then, by further releasing the degree of constriction, one passes through stages which we may call *tense voice*, *voice*, and *lax voice* (though of course recognizing, as with all the stages on this continuum, that there is no predeterminable point at which, for instance, tense voice should be considered voice). Further relaxation leads to a widening of the glottis, particularly between the arytenoids, so that lax voice becomes *murmur*, in which only the anterior portion is vibrating. This state can arbitrarily be distinguished from one in which there is even greater rate of flow through the glottis, which we may now call *breathy voice*. Finally, when even the anterior portion of the glottis is so far apart that it cannot be set in vibration we have the *voiceless* position.”

An example of *lax voice* is the voicing of the vowel onset after the release of the “voiced” Javanese stop (indicated by orthographic *b*) while an example of *tense voice* is the voicing of the vowel onset after the release of the Korean unaspirated fortis stop.

Ladefoged’s 1973 phonation types are about the same as his 1971 ones, except for relatively minor changes. From Halle and Stevens (1971), he adopted *stiff* and *slack* instead of *tense* and *lax* although he uses these terms to indicate distinctions of glottal manner rather than of pitch. He also added the term *spread* to designate an interarytenoid gap more open than *voiceless*, probably reflecting the need to distinguish two voiceless
categories like Catford’s *breath* and *whisper*. The resulting eight phonation types are now:

1) Spread
2) Voiceless
3) Murmur
4) Slack
5) Voice
6) Stiff
7) Creaky
8) Closed

With the exception of *slack* and *stiff*, the remaining phonetic categories are identical to the six degrees of glottal stricture proposed at the beginning of the 1971 article despite changes in terminology.

During the discussion of Laver’s phonatory settings, the question was raised as to the independence of phonation type vis-à-vis pitch register. His classification showed that each voiced phonation type (*plain, harsh, breathy, harsh whispery*) could in principle combine with any of the three pitch registers (*creaky, modal, falsetto*). On the other hand, Ladefoged (1973: 78) points out that production mechanisms may favor some associations over others:

“For example, stiff and creaky voiced sounds (e.g. in Luo and Margi) are produced by moving the arytenoids cartilages closer together. This action necessitates the arytenoids being moved forward. As a result the vibrating parts of the vocal cords are less stretched and these sounds tend to have a lower pitch. This tendency can be counteracted by moving the thyroid cartilage forward and hence stretching the glottis. But unless this additional action is made, whenever stiff or creaky voiced sounds occur the pitch will go down…The opposite kind of effect can be observed in high pitch sounds. These sounds tend to have a breathy
voice quality. When the vocal cords are stretched, the posterior portions (near the arytenoids) will be unable to come together completely during each vibratory cycle. As a result, unless there is a compensating adjustment of the position of the arytenoids, a breathy voice quality will be produced.”

Nevertheless, Ladefoged (1973: 79) cautions against setting up fixed associations in the phonology (e.g. through a cover feature), as glottal strictures can occur on a wide range of pitches. For instance, he puts forward evidence that the stiff or creaky vowels following Korean unaspirated fortis stops show an increase in pitch. He also notes that the murmured sounds in both Indo-Aryan and Bantu are usually realized with a lowering of pitch.

2.3.2 Ladefoged and co-workers (1996, 2001)

There are seven phonation types in the inventory established by Ladefoged and Maddieson (1996: 48, 50, 372):

1) **Voiceless** no vibration of the vocal folds, arytenoid cartilages apart, for a male speaker, volume velocity up to 1000 cm$^3$/s at a subglottal pressure of 8 cm H$_2$O.

2) **Breathy voice (murmur)** vocal folds vibrating but without appreciable contact, arytenoid cartilages further apart, higher rate of airflow than in normal voice, typically 500 cm$^3$/s at 8 cm H$_2$O.

3) **Slack voice** vocal folds vibrating but more loosely, slightly higher rate of airflow than in normal voice, 250 cm$^3$/s at 8 cm H$_2$O.

4) **Modal voice** regular vibrations of the vocal folds at any frequency within the speaker’s normal range, 120 cm$^3$/s at 8 cm H$_2$O.

5) **Stiff voice** vocal folds vibrating but more stiffly than in normal voice, slightly lower rate of airflow than in modal voice, 100 cm$^3$/s at 8 cm H$_2$O.
6) *Creaky voice* (*laryngealized*) vocal folds vibrating anteriorly, but with the arytenoid cartilages pressed together, considerably lower rate of airflow than in normal voice.

7) *Closed* no vibration of the vocal folds, arytenoid cartilages together, no airflow.

Apart from one category of voicelessness instead of two, i.e. *spread* and *voiceless*, and the use of term *breathy voice* instead of *murmur*, the seven proposed phonation types are quite similar to the eight given by Ladefoged in his 1973 study. These general phonetic categories are also the ones presented and discussed in the much more recent article by Gordon and Ladefoged (2001), which provides an extensive bibliography of languages with contrastive nonmodal phonation. As this article treats not only the distinctive function of the phonation types, but focuses on their acoustic correlates as well, a brief summary of its main points will be a good introduction to the final part of this chapter.

*Breathy voice* has been shown to be phonemically contrastive on vowels (Gujarati, Jalapa Mazatec, San Lucas Quiaviní Zapotec) and nonvocalic sonorants (Hindi, Newar, and Tsonga). Breathy phonation is generated by fairly abducted vocal cords with little longitudinal tension and is characterized aerodynamically by turbulent airflow through the glottis, giving an auditory impression of voice mixed with breath. In comparison to normal voice, there is additional spectral noise, particularly at high frequencies. Typically there is a reduction in overall acoustic intensity. The spectral tilt of breathy voice falls off at a significantly faster rate than it does for normal voice. Spectral tilt is often measured by subtracting the logarithmic amplitude of one of the higher harmonics from the logarithmic amplitude of the fundamental; the most frequently used
higher harmonics are the second (H1–H2), the peak harmonic under F1 (H1–A1), or the peak harmonic under F2 (H1–A2). A steep downward spectral tilt is associated with an increased open quotient and a less abrupt glottal closure, a flat spectral slope with just the opposite. Measures of spectral tilt have successfully distinguished between normal and breathy phonation in Jalapa Mazatec, !Xôô, Gujarati, Kedang, Hmong, Tsonga, and San Lucas Quiaviní Zapotec. Breathy voice most often lowers the fundamental frequency, but it is not yet clear how universal the effect is. It also tends to occur with increased pharyngeal width and a depressed larynx, hence with a lower F1 (Chong, Kedang).

*Lax voice* has the same characteristics as breathy voice, e.g. a steep spectral tilt resulting from slack vocal folds, but to a lesser degree (for a similar definition of lax voice, see Ní Chasaide and Gobl, 1997). Mpi and Bruu seem to contrast *lax voice* and *tense voice*.

Much like breathy voice, *creaky voice* enters into phonological oppositions on vowels (Jalapa Mazatec, San Lucas Quiaviní Zapotec), nonvocalic sonorants (Kwakw’ala, Montana Salish, Hupa, Kashaya Pomo), but rarely on obstruents (Chadic implosives). Creaky phonation is produced with tightly adducted vocal folds, but open enough to allow vibration along their length. The resulting acoustic waveform consists of a sequence of irregularly spaced pitch periods, frequently with reduced intensity. There is a considerable increase in jitter (random frequency modulation of the pitch period) relative to the other phonation types (Burmese, Jalapa Mazatec), and the spectral tilt is significantly flatter as well (Burmese, Jalapa Mazatec). The fundamental frequency is usually lowered (Mam, Mohawk, Cayuga, Oneida), but not universally, since the reflexes of a glottal constriction in Proto-Athapaskan are high tones in some languages, but low
tones in others. In addition, the larynx is often elevated, shortening the vocal tract and thus raising F1 (Jalapa Mazatec, Haoni).

*Tense voice* is considered a weaker form of creaky voice. It is produced with stiff vocal folds and thus exhibits a flatter spectral slope than normal voice (see also Ní Chasaide and Gobl, 1997 on tense voice). As noted above, it may be contrastive with lax voice.

This section concludes the survey of the various classifications of phonation type. The following section will present the distinctive phonation types, characterize each of them in terms of acoustic and motor properties, then provide an analysis of the relations existing between them.

2.4 The inventory and organization of the phonation types

2.4.1 The nine phonation types

The nine phonation types illustrated in Table 2.2 appear necessary in order to describe linguistically relevant distinctions in glottal manner. They represent the maximum number of distinctions thus far documented in the world’s languages. The existence and the boundaries of each phonation type must, however, be decided on a language-by-language basis.
As part of their reflections on the phonological categorization of places of articulation, Ladefoged and Maddieson (1996: 4) state:

“As is detailed in our discussion of place of articulation…, this does not mean that boundaries between each category can easily be determined. For example, the tongue has no clearly defined regions, and neither does the roof of the mouth; and there are, of course similar problems in dividing the continua which underlie other phonetic parameters. Nonetheless, it is very striking that languages often seem to cut the continua in similar ways so that it is possible to equate a sound in one language with a similar sound in another. Our procedure in trying to build up a description of the world’s languages is to proceed primarily by a series of such equivalences.”

As part of their reflections on the phonological categorization of places of articulation, Ladefoged and Maddieson (1996: 4) state:

The distinction between Catford’s breath and whisper is a good illustration of the need for a language specific analysis guided by a general model of phonetic categories. In view of establishing a set of glottal features, Gandour (1975) examined the way in which the onset consonants conditioned the three tones of Proto-Tai. His classes of onset consonants are as follows:

I. voiceless aspirated stops and h, symbolized by **ph**

II. voiceless fricatives and voiceless sonorants, symbolized by **f**

III. voiceless unaspirated stops, symbolized by **p**

### Table 2.2 The proposed inventory of nine phonation types organized according to their phonetic characteristics.

<table>
<thead>
<tr>
<th>Phonation Type</th>
<th>Glottal Periodicity</th>
<th>Vocal Fold Position</th>
<th>Spectral Properties</th>
<th>Vocal Fold Stricture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glottal stop Breath</td>
<td>glottal noise</td>
<td>extreme</td>
<td>high Q1</td>
<td>constricted</td>
</tr>
<tr>
<td>Whisper</td>
<td></td>
<td></td>
<td>medium Q1</td>
<td>constricted-spread</td>
</tr>
<tr>
<td>Breath</td>
<td></td>
<td></td>
<td>low Q1</td>
<td></td>
</tr>
<tr>
<td>Harsh voice Breath</td>
<td>noisy voice</td>
<td>central-extreme</td>
<td>low pACC</td>
<td>constricted</td>
</tr>
<tr>
<td>Harsh whispery voice Breath</td>
<td></td>
<td></td>
<td>low pACC&amp;HNR</td>
<td>constricted-spread</td>
</tr>
<tr>
<td>Breathy voice</td>
<td></td>
<td></td>
<td>low HNR</td>
<td></td>
</tr>
<tr>
<td>Tense voice</td>
<td>pure voice</td>
<td>central</td>
<td>flat spectral tilt</td>
<td>constricted</td>
</tr>
<tr>
<td>Plain voice</td>
<td></td>
<td></td>
<td>mid spectral tilt</td>
<td>constricted-spread</td>
</tr>
<tr>
<td>Lax voice</td>
<td></td>
<td></td>
<td>steep spectral tilt</td>
<td>spread</td>
</tr>
</tbody>
</table>

Table 2.2 The proposed inventory of nine phonation types organized according to their phonetic characteristics.
IV. glottal stop, preglottalized stops, and preglottalized glides, symbolized by ?

V. voiced obstruents and sonorants, symbolized by b

These onset classes determine which reflexes of the Proto-Tai tones are to follow, patterning together or separately according to the dialect:

A) White Tai (ph, f, p, ?) vs. (b)

B) Lung Ming (ph, f, p) vs. (?, b)

C) Siamese (ph, f) vs. (p, ?, b)

D) Songkhla (ph, f) vs. (p, ?) vs. (b)

E) Nung (Bac Va) (ph) vs. (f, p, ?) vs. (b)

At the very least, two degrees of open glottis voicelessness must be distinguished in the phonology to account for the grouping of (ph, f) in opposition to (p) in Siamese and Songkhla, i.e. breath vs. whisper, and possibly (ph) to (f, p) in Nung. Gandour goes so far as to hypothesize three degrees (ph = “spread”, f = “voiceless”, p = “tightened”). It seems highly exceptional for the phonetic categories breath and whisper to manifest themselves in a lexically contrastive way. Whether they can be also associated with paralinguistic meanings is a question for future research. Yet the Tai data show that both categories should be integrated into a general framework of phonation types (cf. also Vaux, 1998 for his treatment of Thai tonal reflexes).

The infrequency of contrastive breath and whisper probably led Ladefoged and Maddieson (1996: 48-49) to adopt the traditional cover terms voiceless and voicelessness when referring indifferently to either phonation type, and we will accept these solidly established terms as well. However, they do mention that a glottal stop is technically
speaking also voiceless since it is produced with no vocal fold vibration. To avoid confusion, *glottal stop*, *whisper*, and *breath* will all be considered as manifestations of glottal noise. Remark that White Tai appears to oppose the glottal noise phonations (*ph, f, p, ?*) to the voiced phonation (*b*), an indication that glottal noise may constitute a natural phonological class.

Noncontinuant speech sounds differ from the continuants in that they necessarily evolve over time. Like supraglottal plosives, glottal stops are characterized by three phases: attack, closure, and release. Only one of the three phases, however, may have a directly observable acoustic effect. For example, when the glottal stop is utterance initial and precedes a vowel, only the acoustic transient of the release phase is audible. Similarly, if the glottal stop is utterance final and follows a vowel, then only the attack transient is audible. Between vowels, both attack and release transients are heard, but the glottal closure is silent. On the other hand, the complete glottal closure of an ejective takes on acoustic importance when the vocal tract is excited by the release of the supraglottal stricture.

Because the glottal stop is made with either a complete, or near closure of the glottis (e.g. during the attack and release phases), the damping caused by glottal losses is small. Considerably larger amounts of damping can result from the wider glottal aperture of *whisper*, and a fortiori from the fully open glottis of *breath*. Assuming other parameters remain constant (e.g. the area of the nasal port), an increase in glottal damping leads to a broadening of bandwidth and a consequent reduction in the height of the quality factor Q of the first formant (*Q1*), it being equal to the *F1* resonance frequency (Hz) divided by the 3-dB bandwidth (*B1*) in Hz. The *Q*s of the second and
higher formants fall off at a progressively slower rate because the impedance of the

glottal shunt grows with frequency (Flanagan, 1972: 63-65). Additionally, Stevens (1998:
165-166) writes:

“During normal voicing a typical value for the average glottal opening during the
open phase is about 0.06 cm\(^2\), and this opening would contribute about 120 Hz to
the average first formant bandwidth in this part of the glottal cycle. When the
average glottal opening is greater, as it might be during an aspirated consonant,
the first formant bandwidth is considerably greater. The bandwidth may be so
great that the perturbation method used to calculate it is no longer valid. A glottal
area in excess of 0.2 cm\(^2\) would probably lead to a bandwidth that is comparable
to the frequency of the first formant.”

As Q1 appears fairly sensitive to glottal area, it is a likely indicator of the increasingly
greater glottal openings of glottal noise:

1) **glottal stop** (closed, or very narrow glottal aperture during its attack or release;
high Q1)

2) **whisper** (intermediate glottal aperture; intermediate Q1)

3) **breath** (wide glottal aperture; low Q1)

A method of approximating Q1 will be developed in Chapter 4 and subsequently applied
in Chapter 5.

The next chapter will take a deeper look at the acoustic and motor characteristics
of harshness, breathiness, and the harmonic slope. Consequently, the present discussion is
limited to giving the key acoustic definitions. The noisy voice phonation types *harsh
voice, harsh whispery voice, breathy voice* are similar to those put forward by Laver but,
unlike his, carry no restrictions with respect to pitch register. Any phonation type will be
assumed to combine freely with any pitch level, although there is a tendency for low-
pitched sounds to be harsh, as was noted earlier (on segmentally conditioned pitch and
tone, see Hyman, 1973; Hombert, 1978; Hombert et al., 1979). Following Laver’s
implicit usage, the terms *creaky voice* and *modal voice* will designate only low and mid pitch register respectively, without any implication as to phonation type.

1) *Harsh voice* is defined acoustically as voice with significant modulation noise, i.e. random frequency modulation (jitter) and/or amplitude modulation (shimmer) of the glottal wave. Ladefoged’s terms *laryngealized* and *laryngealization* will be considered interchangeable with *harsh* and *harshness*. Low values of a suitably calculated peak autocorrelation coefficient (pACC) appear to be a good indicator of large amounts of modulation noise.

2) *Breathy voice* is defined acoustically as voice with significant aspiration noise, i.e. broadband noise produced by turbulent aerodynamic flow in the larynx. Low values of the harmonics-to-noise ratio (HNR) are associated with a large amount of aspiration noise.

3) *Harsh whispery voice* is defined acoustically as voice with both significant modulation and aspiration noise. Low values of the pACC and HNR correspond to large amounts of modulation noise and aspiration noise.

The noisy voice phonation types are all characterized by a quasi-periodic glottal source mixed with a significant amount of aperiodicity. The aperiodicity may be inherent in the glottal wave itself (modulation noise) or arise from the additive noise generated by aerodynamic turbulence (aspiration noise). The distinctive use of harsh and breathy voice on vowels and sonorant consonants has been discussed in the preceding survey. In addition, harsh voice is a frequent allophonic variant of the phonemic glottal stop just as breathy voice is often an allophone of voicelessness (cf. the Proto-Tai initial consonant class including both glottal stop and preglottalized glide as well as English intervocalic
/h/ realized as breathy voice; see Gordon and Ladefoged, 2001: 391-392 for further discussion on the allophonic use of noisy voice phonation). As was pointed out in the introduction to this chapter, noisy voice phonation may also signal paralinguistic meaning (cf. Laver’s note on harsh voice to convey anger) or acquire sociolinguistic signification, such as the harsh whispery voice of urban Scots dialects (“whispery ventricular voice”) and the generally breathier voices of women (for the latter, see Hanson et al., 2001 and references therein).

The proposed pure voice phonation types tense voice, plain voice, lax voice are basically those established by Ladefoged and his colleagues as well as by Ní Chasaide and Gobl (1997). Plain voice is used in preference to modal voice, as the latter term, following Laver, indicates mid pitch register and not a phonation type. The pure voice phonation types are interpreted in a more restrictive way than in previous work:

1) Tense voice is defined acoustically as voice with a spectral tilt significantly flatter than plain voice, but with negligible modulation and aspiration noise.

2) Plain voice is defined acoustically as voice with a spectral tilt lying between lax voice and tense voice, but with negligible modulation and aspiration noise.

3) Lax voice is defined acoustically as voice with a spectral tilt significantly steeper than plain voice, but with negligible modulation and aspiration noise.

Like noisy voice, the glottal source of the pure voice phonation types is periodic, but unlike noisy voice, no significant aperiodicity is present. Phonological contrastiveness among the three types of pure voice has been attested in only a few of the world’s languages to date (e.g. Javanese, Korean, Mpi, and Bruu mentioned earlier).
2.4.2 The organization of the phonation types

2.4.2.1 Phonological evidence for a category of glottal noise

The last part of this chapter is devoted to examining how the phonation types are organized among themselves. As was shown in the review, Ladefoged assumes that there is a continuum of glottal opening along which the phonation types are ordered, from the most closed position to the most open. Using the present terms, they are glottal stop, harsh voice, tense voice, plain voice, lax voice, breathy voice, whisper, breath. Observe that it is difficult to place harsh whispery voice on this monotonic continuum.

Nevertheless, one might be justified in disregarding harsh whispery voice since there seems to be no currently available language descriptions in which it is judged phonemically contrastive, contrary to the other phonation types. This question requires further research, however. More serious is the objection that glottal stop and breath are on opposite ends of the continuum, as glottal closure and voicelessness (breath or whisper) can pattern together phonologically.

It has already been pointed out that White Tai opposes the glottal noise phonations (ph, f, p, ?) to voiced phonation (b). In addition, MacEachern (1999: 30-33) gives the following glottal co-occurrence restrictions in Cuzco Quechua:

1) There is only one ejective per morpheme.
2) There is only one aspirated voiceless stop per morpheme.
3) There is only one /h/ per morpheme.
4) Aspirated and ejective stops do not co-occur.
5) Aspirated stops and /h/ do not co-occur.
6) Glottal stop does not co-occur with ejectives.
The voicelessness and glottal closure of co-occurrence restriction 4) appear to be subject to the same kind of dissimulation process as the voicelessness and glottal closure of co-occurrence restrictions 5) and 6), respectively. The pattern of glottal dissimilation found in Cuzco Quechua is not uncommon in other languages that make extensive use of laryngeal contrasts, as MacEachern’s typology of glottal co-occurrence restrictions makes clear. The parallel in behavior suggests that voicelessness and glottal closure are phonetically similar enough—both being glottal noise—that they can be treated alike in phonotactic rules.

Bessell (1992, Chapters 3 and 7) adduces a great deal of evidence indicating that the glottal stop and /h/ often function as the glottal counterparts to voiceless supraglottal stops and fricatives. For instance, she mentions that 38% of the 693 languages in Ruhlen’s 1975 database tend to oppose /ʔ/ and /h/ in this fashion (1992: 40). Moreover, in the standard dialect of Malay, the final stop /k/ is reduced to /ʔ/ while final /s/ optionally becomes /h/. In other dialects the final stops /p, t, k/ are all reduced to /ʔ/ and /s/ obligatorily becomes /h/ (1992: 344). If the glottal stop and /h/ are assumed to belong to the same natural class, i.e. glottal noise, then both weakenings can be considered a unitary phonological process.

2.4.2.2 Typological distribution of the phonation types

The preceding discussion lends support to the notion that the glottal stop and voicelessness (breath or whisper) behave like contrasting subcategories within the single phonological class of glottal noise. However, the single most prevalent glottal distinction is between voicelessness and voice on obstruents. Languages with only two series of
The stops typically make a voiceless-voiced contrast. Of the 162 languages with two series of stops in the 317 language UPSID, 117 (72.2%) have a voicing opposition (Maddieson, 1984: 28-29). When the 76 languages with three series of stops are ranked according to their four most common patterns, it is clear that the voiceless-voiced distinction continues to play an important role:

a) voiceless, voiced, aspirated voiceless
b) voiceless, voiced, ejective
c) voiceless, aspirated voiceless, ejective
d) voiceless, voiced, voiced implosive

All but one c) of the most common patterns of the three stop series shows a voiceless-voiced contrast. Remark also that the glottal stop (ejective) is contrastive in b) and c). Maddieson (1984: 99, 111) considers the default phonation type for the voiced implosive to be laryngealized when not otherwise specified (but see Catford, 1977: 108 and Blust, 1980). Following their frequency of occurrence, the phonation types on the three series of stops are then:

1) Voicelessness (including both voiceless and aspirated voiceless)
2) Voice
3) Glottal stop (ejective)
4) Laryngealized voice (possibly)

When the 25 languages with four series of stops are listed according to their four most common patterns, very similar results are obtained:

a) voiceless, voiced, ejective, voiced implosive
b) voiceless, voiced, aspirated voiceless, voiced ejective
c) voiceless, voiced, prenasalized voiced, voiced implosive or voiced laryngealized

d) voiceless, voiced, aspirated voiceless, breathy voiced

The “voiced ejective” is a prevoiced ejective (see 1984: 100). Therefore the frequency ranking of the phonation types on four series of stops is:

1) Voicelessness

2) Voice

3) Glottal stop (ejective)

4) Laryngealized voice

5) Breathy voice

Because fricatives exhibit a strong tendency to be voiceless, languages usually must have four or more fricatives before a voiceless-voiced contrast begins to manifest itself. Only when the inventory of fricatives reaches eight or more do they become ejective (Maddieson, 1984: 56-57). Ten languages have ejective fricatives, and four more have obstruents variously described as tense, fortis, or voiceless laryngealized (1984: 109-111). Korean, Hausa, and Siona have a tense voiceless alveolar fricative, for example. Thus the phonation type on fricatives is consistent with the stop pattern, but there are relatively fewer kinds of contrast:

1) Voicelessness

2) Voice

3) Glottal stop (ejective)

Sonorants (nasals, liquids, glides, vowels) much prefer voice over voicelessness when compared to obstruents (stops, fricatives). The frequency of voice among nasals
confirms this extreme bias, the number of UPSID nasals indicated by parentheses (1984: 59):

1) Voice (934)
2) Voicelessness (36)
3) Laryngealized voice (34)
4) Breathy voice (3)

The lateral approximant liquids display the same general pattern (1984: 74):

1) Voice (313)
2) Voicelessness (11)
3) Laryngealized voice (8)
4) Breathy voice (1)

As do the trilled rhotic liquids (1984: 78):

1) Voice (130)
2) Voicelessness (3)
3) Laryngealized voice (1)

The frequency distribution of phonation types of the palatal and labiovelar glides are respectively (1984: 92-93, 116):

/j/

1) Voice (273)
2) Laryngealized voice (13)
3) Voicelessness (7)

/w/

1) Voice (240)
2) Laryngealized voice (14)

3) Voicelessness (11)

Lastly, two languages in the database have contrastively laryngealized vowels (Sedang, S. Nambiquara), two languages have a voiceless-voiced contrast on vowels (Ik, Dafla), while only Tamang has contrasting breathy vowels (1984: 132).

Two observations are in order. The primary contrast in phonation type is between voiceless and voiced sounds. This distinction is not only pervasive among stops and fricatives, as expected, but also among nasals and liquids, even though those sonorants are overwhelmingly voiced. Only on glides and vowels does there seem to be a slight trend toward a primary contrast between pure voice and a noisy voice phonation type like laryngealized or harsh voice. However because so few languages are involved, a larger sample is needed to settle this question. The secondary contrast is between phonation types with a strongly constricted glottis and those without a strongly constricted glottis. As shown above, the voiceless obstruents frequently contrast with ejectives as well as voice. In addition, pure voice sonorants are opposed to laryngealized sonorants far more often than they are to breathy sonorants. The primary split between voice and voicelessness is illustrated in Table 2.3.1. The outcome of the secondary split is shown in Table 2.3.2, in which pure voice now contrasts with harsh voice and voicelessness with the glottal stop.
<table>
<thead>
<tr>
<th>Phonation Type</th>
<th>Glottal Periodicity</th>
<th>Vocal Fold Position</th>
<th>Spectral Properties</th>
<th>Vocal Fold Stricture</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Voicelessness</strong></td>
<td>glottal noise</td>
<td>extreme</td>
<td>high Q1, medium Q1, low Q1</td>
<td>constricted, constricted-spread</td>
</tr>
<tr>
<td>Voice</td>
<td>noisy voice</td>
<td>central-extreme</td>
<td>low pACC, low pACC&amp;HNR, low HNR</td>
<td>constricted, constricted-spread, spread</td>
</tr>
<tr>
<td></td>
<td>pure voice</td>
<td>central</td>
<td>flat spectral tilt, mid spectral tilt, steep spectral tilt</td>
<td>constricted, constricted-spread, spread</td>
</tr>
</tbody>
</table>

**Table 2.3.1** The primary contrast in phonation type between voicelessness and voice according to UPSID frequencies.

<table>
<thead>
<tr>
<th>Phonation Type</th>
<th>Glottal Periodicity</th>
<th>Vocal Fold Position</th>
<th>Spectral Properties</th>
<th>Vocal Fold Stricture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glottal stop</td>
<td>glottal noise</td>
<td>extreme</td>
<td>high Q1, medium Q1, low Q1</td>
<td>constricted, constricted-spread</td>
</tr>
<tr>
<td>Voicelessness</td>
<td>noisy voice</td>
<td>central-extreme</td>
<td>low pACC, low pACC&amp;HNR, low HNR</td>
<td>constricted, constricted-spread, spread</td>
</tr>
<tr>
<td></td>
<td>pure voice</td>
<td>central</td>
<td>flat spectral tilt, mid spectral tilt, steep spectral tilt</td>
<td>constricted, constricted-spread, spread</td>
</tr>
</tbody>
</table>

**Table 2.3.2** The secondary contrast in phonation types with strongly constricted glottis and without strongly constricted glottis according to UPSID frequencies.
2.4.2.3 Motor evidence for a hierarchy of phonation type

It is of interest to examine more closely the motor and acoustic properties underlying these successive divisions. As will be discussed in the next chapter, the greater the degree of periodicity of vocal fold vibration, the higher the value of the peak autocorrelation coefficient (pACC). In principle, each language sets a fuzzy boundary bisecting the continuum of pACC values into the two primary categories glottal noise and voice, but the fixed threshold experimentally determined in Chapter 4 appears to work fairly well across languages.

A further bisection of voice gives rise to the secondary contrast between harsh voice and pure voice, thereby establishing a three-way partition of the pACC values: glottal noise, noisy voice, and pure voice. Likewise, the bisection of glottal noise yields the secondary contrast between voicelessness and glottal stop.

Acoustically, the primary contrast between glottal noise and voice is cued by the peak autocorrelation coefficient (pACC). Motorically, the secondary contrast between the presence and absence of a strongly constricted glottis is due to the sphincteric contraction of the laryngeal adductor musculature and its relaxation (e.g. the lateral cricoarytenoid and interarytenoid muscles, the latter made up of the transverse arytenoid, oblique arytenoid, and aryepiglottic; Zemlin, 1981: 163-166).

The question can be asked as to what is the motoric manifestation of the primary contrast between glottal noise and voice. One of the aerodynamic conditions for vibration of the vocal folds to occur is that they must sufficiently approach each other along the center line, yet not be excessively pressed together so as to impede movement.
Stevens (1998: 80) states that:

“We consider first a range of conditions for which the vocal folds are adducted, so that in the rest position the vocal folds are pushed together. Vocal fold vibration will occur only if the transglottal pressure provides sufficient force to push the vocal folds apart. Otherwise, the glottis remains closed, and there is no flow of air.

Vocal fold vibration can also fail to occur when the glottis is sufficiently abducted. When the alveolar pressure is applied, the energy supplied by aerodynamic forces in a cycle of vibration is not sufficient to overcome energy losses in the vocal folds, and hence vibration is not maintained.”

Thus a central position of the vocal folds is critical to the initiation and maintaining of voice. On the other hand, extreme positions of the vocal folds characterize glottal noise. The vocal folds are tightly constricted during the glottal stop, but spread wide open for breath, whereas whisper combines the constricted glottis of the former and the spread glottis of the latter. Because whisper represents an apparently contradictory glottal state it merits closer examination.

Monoson (1976: 4-6) reviewed the descriptions of the glottal configuration of whisper put forward by several authors. There seems to be a consensus that the anterior or ligamental glottis is adducted whereas the posterior or cartilaginous glottis is abducted, the latter forming a triangular aperture through which air is forced. However, there can be varying degrees to which the anterior glottis is adducted and the posterior glottis is abducted. Griesman (1943) first observes that “the larynx can be closed at three levels: at the level of the true vocal cords, at the level of the ventricular folds, and at the level of the aditus laryngis, the aryepiglottic sphincter” (see Pressman, 1954 for further discussion on the three levels of glottal closure). He then goes on to describe the results of X-ray planigraphy of whisper during the low vowel of “father”:

“It is known through laryngoscopic examination that a wide triangular gap is often present posteriorly during whispering. But such examination does not seem
to show the real appearance of the larynx during whispering. No pyriform sinus or ventricle is visible. The ventricle does not give free play to the vibrating cords. One can see the narrow subglottic space as a sign of decreased air pressure, the dome-shaped inferior aspect of the approximated vocal cords, with a clearly blurred gap, indicating that the vocal cords perform real vibrations. The supraglottic space is obliterated because of approximation of the ventricular folds and the aryepiglottic folds above them. There is only a small outlet for the escaping air. The sphincter girdle is contracted at all three levels, but this contraction is less pronounced than in the contraction of the same girdle for fixation of the thorax.”

Observe the similarity of this tightly constricted kind of whisper with the glottal configuration of the pharyngeal fricative [h], as described by Esling (1999a, see also 1996):

“Viewed laryngoscopically, both [h] and [?] demonstrate aryepiglottic fold constriction. For a voiceless fricative, a medial triangular opening remains between the arytenoids, above the level of the glottis, as they press forward against the epiglottic tubercle. This funneling of the laryngeal vestibule can be hypothesized to be the source of friction giving rise to voicelessness.”

In an X-ray study of Arabic, Ghazeli (1977: 49) considers the strong whisper of [h] to be an essential element in its production:

“My conviction is that the constriction of the pharynx during the production of the pharyngeals also results in the constriction of the larynx. This results in a somewhat creaky voice noticeable during the articulation of the voiced pharyngeal [?] During [h] this constriction may result in a special state of the glottis where the vocal cords are either tense or only open in one end. The constricted larynx may be used to rush the air against the root of the tongue to create the high intensity noise perceived during the production of [h]. Although this is highly speculative, it seems to me that the friction during [h] is created by interlaryngeal adjustments rather than the relatively wide constriction between the epiglottis and the pharyngeal wall.”

Similarly to strong whisper, the strong glottal stop or epiglottal plosive [?] appears to be produced with three levels of glottal closure (true vocal folds, ventricular folds,
aryepiglottic sphincter). Esling (1999a) describes the sound in the following way (references omitted):

“Hockett identified a ‘pharyngeal catch’ (as distinct from a continuant) in some dialects of Arabic, and Catford has noted a ‘pharyngealized glottal stop’ or ‘strong glottal stop’ in the languages of the Caucasus, and a pharyngeal stop in Chechen. Butcher and Ahmed observed that the voiced pharyngeal in Iraqi Arabic sometimes functions as a stop. Early work by Stephen Jones at University College London influenced Catford to draw a relationship between pharyngeal constriction and the ventricular phonatory setting, which occurs when ‘the ventricular bands are brought together…, plus some generalized constriction of the upper larynx and pharynx,’ so that ‘ventricular or strong glottal stop may be represented by [ʃˀ],’ in contrast to [ʔ]. Gaprindashvili described this articulation as a ‘pharyngealized glottal stop’. The strong glottal stop occurs in the Nakh languages and some Daghestanian languages, and is sometimes termed a ‘pharyngeal stop’ in the Georgian literature. The sound has been isolated as an ‘epiglottal plosive’ in the inventory of the IPA, and given the symbol [ʔ]. The location of stricture has been identified at the aryepiglottic folds—at the laryngeal sphincter mechanism, the third level of closure above the glottis and the ventricular folds—which seals the airway in an anterior movement against a descending tongue root. It appears from this research that the larynx usually rises to lift the arytenoids forward to effect the most efficient strongest seal.”

It is evident from the preceding accounts that whisper and the glottal stop share the common motor property of a relatively tightly constricted glottis. On the other hand, whisper resembles breath in that both are characterized by airflow through a spread glottal opening. Whisper thus appears to be a motor compromise between the extremely adducted glottal stop and the extremely abducted phonation type of breath. The joint occurrence of the two strongly antagonistic motor actions of whisper is indicated by the term constricted-spread in Tables 2.2, 2.3.1, and 2.3.2.

As pointed out previously, the vocal folds must be approached in the center line for voice production to take place. Once a central phonatory setting is established for pure voice, for example, the vocal folds may then be slightly adducted to produce tense voice or slightly abducted to yield lax voice. Plain voice occurs when the two weakly
antagonistic motor actions are implemented simultaneously, giving rise to an average glottal width midway between tense voice and lax voice. Note the resemblance with the middle portion of Ladefoged’s continuum of glottal stricture:

1) Tense voice
2) Plain voice
3) Lax voice.

When the vocal folds adopt a position that is both central and extreme, the resulting sounds combine pure voice and glottal noise, hence the term noisy voice for the three central-extreme phonation types (harsh voice, harsh whispery voice, breathy voice). In harsh voice, the vocal folds are central enough to allow vibration and are thus more moderately adducted than the glottal stop. However, some realizations of harsh voice can approach the extreme adduction of the strong glottal stop. The intense squeezing of the ventricular folds during severe harsh voice—according to Laver’s description—is a good illustration of this. In breathy voice, the vocal folds are again central enough to vibrate and thus are more moderately abducted than during breath. Harsh whispery voice, in a manner highly reminiscent of whisper, represents the motor compromise between moderately constricted harsh voice and moderately spread breathy voice, but unlike whisper, the vocal folds are positioned to be sufficiently central so that voice can occur.

Summarizing, there appears to be some support for considering the maximum inventory of phonation types to be arranged motorically in a three-by-three fashion. The primary motor dimension of position is characterized by opposing action between an extreme vocal fold position (glottal noise) and a central one (pure voice), the central-extreme position (noisy voice) resulting from the simultaneous implementation of the two
polar positions. The secondary motor dimension of stricture involves antagonistic action between a constricted vocal fold stricture and a spread one, thereby leading to a maximum of three degrees of stricture (constricted, constricted-spread, spread) within each of the three positions established by the primary motor dimension (see Figure 2.1). The constricted-spread gesture arises from the joint implementation of constricted and spread vocal folds as, for example, during *plain voice, harsh whispery voice, whisper*.

There appears to be no electromyographic (EMG) study that systematically explores the different phonation types using the same participants and procedures. Nevertheless, a sufficient number of studies are available to evaluate the plausibility of the proposed schema. The interarytenoid adductor muscle (INT) regularly shows activity during voicing, as one would expect since the vocal folds must be rather closely approximated in order to vibrate (Faaborg-Anderson, 1957; Hiroto et al., 1967; Hirose and Gay, 1972; Sawashima and Hirose, 1983). The role of the posterior cricoarytenoid abductor muscle (PCA) has been less clear, with early studies showing little activity during voicing (Faaborg-Anderson, 1957; Hiroto et al., 1967) and the more recent ones demonstrating greater activity, particularly at higher pitch ranges (Gay et al., 1972; Hirano, 1987). In their investigation of canine PCA muscles, Choi et al. (1993) concluded “our data suggest that the PCA muscle not only braces the arytenoids against anterior pull, but also serves as an antagonist by regulating phonatory glottal width.” The central position of the vocal folds during normal voicing therefore appears to be the result of a relatively strong INT adductive component and a weak PCA abductive component acting as its antagonist. Hence *plain voice* is not only central, but very likely constricted-spread.
Figure 2.1 The nine phonation types arranged according to the three vocal fold positions of the primary motor dimension (central, central-extreme, extreme) as well as by the three vocal fold strictures of the secondary motor dimension (constricted, constricted-spread, spread) within each of the positions of the primary motor dimension.
as well. Hirose and Gay (1972) observed partial activation of the PCA and slight suppression of the INT during poststressed voiced fricative /z/ and believed the effect to be related to the often greater glottal opening of voiced fricatives. This apparent instance of lax voice illustrates how the respective neural inputs to the PCA and INT could be weighted to achieve a central-extreme setting like breathy voice.

The PCA, generally considered the sole abductor, swings the vocal folds to the extreme spread position of breath when the INT activation level is low (see Hirose and Gay, 1972 on the reciprocal activity of the PCA and INT). Löfqvist and Yoshioka (1981) show that the Swedish voiceless fricative and postaspirated stop exhibit about the same PCA peak values. They also remark that in Icelandic:

“…the postaspirated stops have a larger glottal opening than their preaspirated and unaspirated cognates. Glottal opening is smaller for the preaspirated type, and very small for the unaspirated one. For the latter, the fiberoptic films showed a small, spindle-shaped opening in the membranous portion of the glottis.”

Gay et al. (1972) noted a marked increase in the activity of the lateral cricoarytenoid adductor (LCA) before a glottal attack, considering it to be an indicator of strong medial compression. Similarly high activity of the LCA was obtained during voluntary cough (Poletto et al., 2004). Thus, the extreme constricted glottal stop is characterized by significant activity of the LCA, most likely in synergy with the INT adductor. Not surprisingly, in view of the earlier discussion, whisper is also associated with significant LCA activation, together with an increase in PCA abductor activity and a decrease in INT adductor activity (Sawashima and Hirose, 1983; see also Weitzman et al., 1976). Despite the dearth of EMG studies dedicated specifically to phonation types, this rapid overview shows that the EMG evidence is in harmony with observations gained by more direct methods, and is not incompatible with the proposed motor mechanisms.
2.4.2.4 Evidence for a linear acoustic scale of phonation type

Acoustic periodicity generally increases as the vocal folds assume a more central position on the primary motor dimension of position: extreme (glottal noise), central-extreme (noisy voice), central (pure voice). As will be discussed in the next chapter, the peak autocorrelation coefficient (pACC) provides an excellent diagnostic of acoustic periodicity. Upon inspection of Table 2.2, however, it becomes apparent that the opposition between constricted and spread vocal folds on the secondary motor dimension of stricture is not so easily quantified by a single acoustic measure. For example, a high Q1 signals the constricted glottis of the *glottal stop* and a low Q1 the spread glottis of *breath*. A small pACC (significant modulation noise) indicates the constricted glottis of *harsh voice* and a small HNR (significant aspiration noise) the spread glottis of *breathy voice*. A flat harmonic slope cues the constricted glottis of *tense voice* and a steep harmonic slope the spread glottis of *lax voice*. At first glance then, the major acoustic cues for the distinction in glottal stricture seem to have very little in common.

Note, however, that *harsh voice* has a flat harmonic slope like *tense voice*, in addition to significant modulation noise. Similarly, *breathy voice* has a steep harmonic slope like *lax voice*, together with significant aspiration noise. As a result, the minor acoustic cues for *harsh voice* and *breathy voice* are identical to the major acoustic cues for *tense voice* and *lax voice*, respectively. Thus, whether they belong to pure or noisy voice, the constricted phonation types are characterized by a flat spectrum with a large overall glottal bandwidth, the spread phonation types by a steeply falling spectrum with a smaller overall glottal bandwidth. The voiced phonation types are not alone in using overall glottal bandwidth to signal changes in vocal fold stricture. The glottal noise
phonation types do likewise. For example, the constricted glottis of the glottal stop displays a nearly flat spectrum because of the unattenuated first formant (high Q1). The shunt losses incurred by the spread glottis of breath, on the other hand, decrease the amplitude of the first formant to a considerable extent (low Q1), effectively reducing the overall glottal bandwidth. Hence glottal bandwidth serves as a derived common cue for changes in the constricted vs. spread dimension over all three degrees of acoustic periodicity (glottal noise, noisy voice, pure voice).

In a highly influential article on language development, Jakobson (orig. 1941, 1971: 375; see Ingram, 1992 for references and discussion) posited that the first phonological opposition to be acquired is between the closed labial stop and the optimally open vowel /a/, the simplest and maximal contrast (“dieser einfachste und maximale Kontrast”) available to the infant. Because the release of a labial plosive is pulse-like and characterized by a flat spectrum envelope, the overall bandwidth is very large, up to 10 kHz (Fant, 1960: 192). On the other hand, the first and second formants of a vowel, the ones most important for vowel quality, seldom exceed 3 kHz. Thus acoustically, the maximal contrast is between a broadband consonant and a relatively narrow-band vowel. Taking this notion to the extreme limit, the optimal consonant would then consist of a pulse with infinite bandwidth, the optimal vowel a pure tone with infinitesimal bandwidth.
Not only can changes along the constricted vs. spread dimension affect overall glottal bandwidth, but changes in glottal periodicity can modify it as well. For instance, glottal noise, as exemplified by the pulse-like attack or release of the *glottal stop*, is characterized by a very broad bandwidth. By contrast, pure voice, as exemplified by *lax voice*, may come to resemble a pure tone when the harmonic slope is exceedingly steep, as in certain varieties of falsetto (Laver, 1980: 120). Nevertheless, the basic cue for glottal periodicity almost certainly remains the neural analog of the pACC because it is utilized to estimate F0 as well. The glottal bandwidth of noisy voice most likely lies between glottal noise and pure voice as it shares the wide-band and narrow-band characteristics of both. Consequently, the derived glottal bandwidth is assumed to decrease monotonically from phonation type 1 to 9, as indicated in the GBW column of Table 2.4.

This arrangement of glottal bandwidth values may help explain why contrasts between phonation types with strongly constricted glottis and those without strongly constricted glottis occur so often in the UPSID. Recall that these contrasts rank second in

Table 2.4: The nine phonation types arranged 1 through 9 according to decreasing broad-spectrum derived glottal bandwidth GBW.

<table>
<thead>
<tr>
<th>GBW</th>
<th>Phonation Type</th>
<th>Glottal Periodicity</th>
<th>Vocal Fold Position</th>
<th>Glottal Stricture Bandwidth</th>
<th>Vocal Fold Stricture</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Glottal stop</td>
<td>large</td>
<td>constricted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Whisper</td>
<td></td>
<td>extreme</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Breath</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Harsh voice</td>
<td>noisy voice</td>
<td>central-extreme</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Harsh whispery voice</td>
<td>noisy voice</td>
<td>central-extreme</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Breathy voice</td>
<td>pure voice</td>
<td>central</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Tense voice</td>
<td>pure voice</td>
<td>central</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Plain voice</td>
<td>pure voice</td>
<td>central</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Lax voice</td>
<td>pure voice</td>
<td>central</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.4: The nine phonation types arranged 1 through 9 according to decreasing broad-spectrum derived glottal bandwidth GBW.
frequency relative to the primary distinction opposing voice to voicelessness, and that the contrasts involve the two sets of oppositions: 1) voicelessness vs. glottal stop, 2) pure voice vs. harsh voice. To maintain distinctiveness, one would expect phonation types to be kept apart as much as possible in terms of glottal bandwidth, and this appears to be the case. For example, the distance between average pure voice and harsh voice \[|\text{GBW}(8) - \text{GBW}(4)| = 4\] is greater than the one separating average pure voice and breathy voice \[|\text{GBW}(8) - \text{GBW}(6)| = 2\]. The preference for maximal dispersion along a linear acoustic scale of glottal bandwidth may account for the fact that contrastive harsh voice is far more frequent than breathy voice in the UPSID. On a motor level, it is not surprising that a central-extreme phonation type (noisy voice) should be chosen to contrast with a central phonation type (pure voice) as both are the result of the second bisection of voice into two opposing categories: central-extreme vs. central. However, there is no apparent reason why central-extreme harsh voice should be overwhelmingly preferred to central-extreme breathy voice unless the constricted glottis is selected 1) by analogy to the voicelessness vs. glottal stop contrast and/or 2) in order to enhance the difference in glottal bandwidth.

It seems likely then that there are two concurrent ways in which the phonation types are organized. One is through the interplay of antagonistic actions in the motor domain, leading to a three-by-three hierarchy of phonation types:

I. Primary motor dimension of position: extreme ~ central-extreme ~ central

II. Secondary motor dimension of constriction: constricted ~ constricted-spread ~ spread.
The other is through a linear scale of derived glottal bandwidth (GBW) in the acoustic
domain, along which the phonation types are ordered. This question will be further
discussed in Chapter 7 where a set of glottal manner features is proposed. The main focus
of the next chapter will be to review the acoustic characteristics of aspiration noise,
modulation noise, and the spectral tilt as well as the various methods of estimating each.
CHAPTER 3

ACOUSTIC PARAMETERS OF PHONATION TYPE

3.1 Introduction

The association between breathy voice and aspiration noise noted in Chapter 2 has led researchers to devise numerous methods for measuring the amount of additive noise in the speech signal (see Buder, 2000 for a comprehensive survey up to 1990). Computed in either the time or the frequency domain, the measures attempt to separate the quasi-periodic from the random turbulent components of the glottal wave. Nearly all these techniques were designed with the goal of achieving objective assessment of disordered voice. Little consideration has been given to their potential auditory neural implementation (but see Shrivastav, 2001). While breathy voice is a combination of pure voice and aspiration noise, harsh or laryngealized voice is associated with irregular pitch periods or modulation noise. These irregularities are manifested by frequency (jitter) and amplitude (shimmer) modulations of the glottal wave.

Spectral tilt is another parameter that has been shown to vary with phonological breathiness and harshness. Fischer-Jørgensen (1967) developed several spectral tilt measures to distinguish between the modal and breathy vowels of Gujarati. One of these, $H_{1-H2}$, is the amplitude difference in dB between the first and second harmonics. With the exception of one speaker, the means of $H_{1-H2}$ were greater for the breathy vowels than they were for the nonbreathy vowels. In the same study higher frequency tilts were estimated by the measures $H_{1-A2}$ and $H_{1-A3}$, each of which representing the amplitude
difference between the first harmonic and the peak harmonic in the second and third formant ranges. The measure H1–A2 yielded fewer classification errors than H1–A3.

In sum, aspiration and modulation noise are likely candidates as cues for phonologically breathy and harsh voice, respectively. Measures of spectral tilt have been used as cues for them as well. The remainder of the chapter is dedicated to a closer examination of these acoustic measures as well as their motor correlates.

3.2 Laryngeal Noise

3.2.1 Mechanisms of modulation noise

Titze (1994: 283-288) details four possible origins of perturbations in the glottal waveform:

1) Neurologic: The movements of the vocal folds and the respiratory system depend on large ensembles of neurons with stochastic firing intervals. The resulting neural noise contributes to the variability of the glottal waveform.

2) Biomechanic: Differences in the biomechanical characteristics of the two vocal folds can desynchronize the vibratory pattern, even in excised larynges devoid of neural activity.

3) Aerodynamic: High rates of airflow through the glottal opening give rise to jet turbulence.

4) Acoustic: Coupling between the vocal tract and the folds can affect the vibratory pattern to a certain extent.

Of the four perturbing influences discussed by Titze, only the aerodynamic and biomechanic are subject to linguistic control. The speaker can regulate the airflow and
adjust the aperture of the vocal folds so as to produce aerodynamically generated aspiration noise. The vocal folds can also be configured in such a way that the biomechanical characteristics of the vibration are altered, resulting in period-to-period irregularities or modulation noise. By contrast, the effects of neural noise and obligatory acoustic coupling lie outside the domain of linguistic intention.

The early work of Lieberman (1961) sought to establish a baseline for pitch perturbations during normal speech. The absolute differences between the durations of hand-measured adjacent pitch periods were compared with their normalized frequency of occurrence. The differences were greater than 0.6 ms 20% of the time and grew with increasing period durations up to 6 ms. However large jumps in the pitch period (pitch breaks) often occurred at the start or end of voicing and during rapid spectral changes, possibly indicating strongly nonlinear behavior of the vocal fold system.

Current theoretical perspectives on the nonlinear dynamics of vocal fold vibration took shape mainly in the first half of the 1990s (Titze et al., 1993; Herzel et al., 1994; see Herzel, 1993 for a historical overview). Since then, several algorithms modeling the different aspects of nonlinear vocal dynamics have been developed, complemented by direct and indirect laryngeal observation. Nevertheless, an understanding of the many complexities of nonlinear vocal dynamics is still very much a work in progress. In the following account, the nonlinear behavior of the vocal folds will be discussed in view of potential cues for modulation noise.

Before considering studies devoted specifically to nonlinear vocal fold vibration, it is useful to examine the somewhat more tractable problem of a forced system consisting of a mass attached to a rigidly fixed spring with linear viscous damping.¹
Assuming cubic nonlinearity of the spring and a sinusoidal excitation function $A \cos \omega t$, the resulting equation of motion is known as the Duffing equation

$$\ddot{x} + \omega_0^2 x + 2\epsilon \mu \dot{x} + \epsilon \alpha x^3 = A \cos \omega t,$$

(3.1)

where $x$, $\dot{x}$, and $\ddot{x}$ are respectively the displacement and its first and second time derivatives, $\epsilon$ is a small constant, $\mu$ a positive damping parameter, $\omega_0$ the angular resonance frequency, $A$ the excitation amplitude, and $\omega$ the angular excitation frequency.

A linear spring ($\alpha = 0$) displays a proportional stress-strain relation. A positive value of $\alpha$ represents a hardening nonlinear spring that requires ever more stress to increase strain while a negative value of $\alpha$ indicates a softening nonlinear spring. The solution of Equation 3.1 consists of two parts: a free oscillation term and a steady state response term. In the free oscillation response of a linear system ($\alpha = 0$), the system parameters fix

**Figure 3.1** Resonance curves for (a) a linear spring, (b) a hardening nonlinear spring, (c) a softening nonlinear spring (adapted from Nayfeh and Mook, 1979).
the natural frequency and the rate of decay, while the initial conditions determine the amplitude and phase. The steady state response depends on the damping parameter, the resonance frequency, the excitation amplitude, and the excitation frequency. As shown in (a) of Figure 3.1, the linear steady state response displays a central peak at $\omega_0$ with the bandwidth of the resonance curve specified by the damping parameter. For a hardening nonlinear spring ($\alpha > 0$), the resonance curve (b) is bent increasingly toward the right as the excitation amplitude becomes larger and the damping parameter smaller.

Consequently, there may be more than one response for each $\omega$, leading to a pattern of sharp discontinuities. When $\omega$ is gradually lowered below point A, the magnitude of the response will follow the smooth curve AFB. As soon as point B is reached, however, any further lowering will induce a jump up to point C. The response then declines along the curve CD. When $\omega$ is raised in the opposite direction along the curve DCE, the response passes through C all the way up to E, at which point it falls abruptly down to F. The discontinuities are reversed for the softening nonlinear spring ($\alpha < 0$) where the resonance curve (c) bends to the left. Thus when $\omega$ varies over the three regions BF, BE, and CE, only the first and last allow for a periodic steady state response, BE being an unstable saddle point. The initial conditions determine which of the two stable periodic solutions occurs in contrast to linear systems that are independent of initial conditions once the free oscillation has decayed to zero. In addition to jump phenomena, there is another consequence of nonlinearity: the generation of additional frequency components. If the damping parameter is small and the excitation amplitude sufficiently large for the hardening nonlinear spring, the free oscillation response may show no signs of decay at higher excitation frequencies. Instead of being fixed by the system parameters as in the
linear case, the frequency of the free oscillation takes on a value 1/3 that of the excitation frequency. This phenomenon is known as subharmonic resonance (Nayfeh and Mook, 1979: 181-182).

Using analog and digital computers, Ueda (1980) investigated the behavior of a simplified Duffing equation without the linear $\omega_0^2 x$ term

$$\ddot{x} + k \dot{x} + x^3 = B \cos t,$$

(3.2)

where $k$ and $B$ are the nondimensionalized damping parameter and excitation amplitude, respectively. When $k$ and $B$ is varied as well as the initial conditions $\left(0 \leq B \leq 25, 0 \leq k \leq 0.8\right)$, a large number of different stable periodic responses (limit cycles) are observed. However, as in the previous example, small changes in the parameters and the initial conditions often lead to jump phenomena. Subharmonic and superharmonic resonances of order $m/n$ also occur, where the response frequency is $m/n$ times the excitation frequency ($m = 1, 3, 4, 5, 6, 7, 11$ and $n = 2, 3$). Furthermore, the superharmonic resonance of order 7/3 in the left region of the $Bk$ plane $\left(6 \leq B \leq 11, 0 \leq k \leq 0.35\right)$ is flanked on both sides by either chaotic motion or chaotic motion alternating with periodic motion depending on the initial conditions. Chaotic motion is nonperiodic motion, but it is nevertheless deterministic because its time evolution is governed by a nonlinear differential equation. The waveform is characterized by long-term unpredictability as well as pulse-to-pulse irregularities in period and amplitude. A chaotic waveform is produced by Equation 3.2 when $B = 12$ and $k = 0.1$ (Thompson and Stewart, 1986: 8). Its power spectrum is a falling line spectrum at the excitation frequency and its odd multiples, interspaced with broad regions of continuous
spectrum noise (Ueda, 1979). This broadband noise is generated uniquely by the equation of motion, and thus has a different origin from highly uncorrelated random noise.

In an early one-mass approximation of the vocal fold system, each fold was represented by a mass attached to a rigidly fixed spring and a nonlinear damping parameter that increased stepwise from relatively low to critical damping on vocal fold contact (Flanagan and Landgraf, 1968). The entire system consists of opposing masses driven by subglottal and Bernoulli pressures. Although the control parameters could be tuned to produce plausible glottal waveforms, the model showed excessive acoustic interaction between the source and the vocal tract. More realistic two-mass models were developed with each vocal fold being approximated by two superimposed masses coupled with a linear spring (Dudgeon, 1970; Ishizaka and Flanagan, 1972). In the Ishizaka and Flanagan model, the damping characteristics of the coupled masses are similar to those of the one-mass system, whereas the fixed springs are given a combination of linear and nonlinear properties varying as a function of vocal fold displacement, glottal rest area, and vocal fold contact. The results of their model are consistent with a number of characteristics of normal phonatory behavior, including relative independence from vocal tract shape. Had the parameter space been fully explored, it is likely that such behaviors as jump phenomena, subharmonic resonances, and chaotic motion would also have been observed given the nonlinear nature of the damping and stiffness parameters.

To address the question of modulation noise, Ishizaka and Isshiki (1976) incorporated into the Ishizaka and Flanagan model an additional nonlinear hardening spring between the colliding masses. The parameters of each of the opposing vocal folds
were also varied independently so as to create asymmetries between them. There are two relevant findings in the study:

1) Irregular motion with period doubling (pitch halving due to a subharmonic resonance of order 1/2) occur at the normal glottal rest area of 0.05 cm² when a) the subglottal pressure is increased to 16 cm H₂O and b) the stiffness of one of the bottom springs is set somewhat lower than its Ishizaka and Flanagan value. However at 20 cm H₂O, the motion nearly regains its primary periodicity under the same area and stiffness conditions. This discontinuous response to a monotonic increase in forcing amplitude is similar to that observed for the Duffing equation.

2) Irregular motion with period doubling and tripling (subharmonic resonances of order 1/2 and 1/3) occurs at the normal subglottal pressure of 8 cm H₂O when a) the glottal rest area is narrowed (< 0.05 cm²) and b) the stiffness of one of the bottom springs is small compared to its Ishizaka and Flanagan value. Note that this description corresponds rather well to the phonation type of harsh creaky voice, which is characterized by irregular periodicity as well as by adducted and relatively slack vocal folds.

A more recent study of asymmetrical vocal fold behavior by Steinecke and Herzel (1995) uses a simplified Ishizaka and Flanagan model to examine the transition or bifurcation points to subharmonic resonance and chaotic motion. Two parameters are adjusted, the subglottal pressure and the tension imbalance between the two vocal folds. When the subglottal pressure is kept fairly high and the tension imbalance varied, several types of bifurcations take place among which: 1) period doubling consisting of cycles with two
pulses alternating in amplitude, both folds being phase-locked in a 2:2 ratio (cf. the 1:1 entrainment pattern of nonnoisy phonation), 2) biphonation or two incommensurate fundamental frequencies caused by a 5:8 entrainment of vocal fold modes (one fold with five pulses per period, the other with eight), 3) irregular motion with the fundamental and the first subharmonic immersed in broadband chaotic noise. Overall, the study indicates that subharmonic regimes of low order are quite frequent while chaotic motion is restricted to a small region of the parameter plane (see also Mergell et al., 2000). Berry et al. (1996) investigated the phonatory behavior of five excised canine larynges. Not only did adduction and elongation asymmetries lead to the expected subharmonic regimes, but one larynx with normal vocal folds evidenced period doubling at the fairly high subglottal pressure of 16 cm H₂O.

Although the vibration pattern displays particular sensitivity to vocal fold asymmetry, there is ample evidence that nonlinear behavior can be induced in the symmetrical folds as well. Herzel and Knudsen (1995) set the stiffnesses of the upper springs to zero while retaining the coupling springs in a symmetrical Ishizaka and Flanagan model. They observed a rather complex glottal waveform for a small value of the coupling stiffness, but a normal waveform for an incrementally larger one. Modeling the symmetrical vocal folds using a more realistic multi-mass finite-element approach, Berry et al. (1994) found that a lax cover gives rise to subharmonic and chaotic regimes. When the transverse Young’s modulus of the cover fell from its initial value of 2 kPa to 0.6 kPa, the vibratory pattern shifted. There was irregular motion mixed with a periodic component between 0.58 and 0.42 kPa, then at 0.4 kPa a period doubling took place. The irregular motion was attributed to the desynchronization of the lower modes of vocal fold
vibration. Neubauer et al. (2001) showed that irregularities along the anterior-posterior dimension of the symmetrical vocal folds could lead to biphonation. They proposed that nonnoisy phonation could be accounted for by the first two modes of the oscillating vocal folds in a symmetrical system, biphonation by the first three modes in an asymmetrical system, and biphonation by the higher-order modes in a symmetrical system with anterior-posterior nonuniformity. Švec et al. (1996) studied the normal vocal folds of a male subject by means of electroglottography and photoglottography. A period doubling bifurcation occurred when the vocal folds were adducted and the airflow was simultaneously increased. Finally, Jiang et al. (2001) used the simplified Ishizaka and Flanagan model of Steinecke and Herzel (1995) to examine the motion of the symmetrical vocal folds. At a given subglottal pressure, a reduction in glottal rest area exerted a strong threshold-lowering effect on the appearance of irregular vibrations. However, for system parameters other than glottal rest area, the relation was nonmonotonic. While very high subglottal pressure consistently led to chaotic noise in the vocal fold spectrum, there were jumps in the response when the stiffness was varied at a lower subglottal pressure (cf. the Ishizaka and Isshiki study for similar behavior in an asymmetrical model).

To summarize, four motor mechanisms contribute to the likelihood that the nonlinear vocal fold system will exhibit pitch breaks, biphonation, subharmonic regimes, and irregular vibration:

1) adduction and elongation asymmetries between the two vocal folds
2) greater than normal adduction of the vocal folds (small glottal rest area)
3) nonuniform vocal fold parameters, particularly those involving a tension decrease
4) heightened subglottal pressure

Assuming normal symmetrical vocal folds (thereby excluding the first mechanism), it is useful to consider how the three remaining factors are related to harsh voice, the phonation type associated with irregular vibration or modulation noise. As was discussed in Chapter 2, creaky voice is low pitch register with no implication as to the presence of irregular vibrations. If significant period-to-period irregularity is present, then low pitch register voice is properly termed harsh creaky voice or laryngealized creaky voice (see Table 2.1). Creaky voice is irregular so frequently (see Michel, 1964), however, it is easy to understand why this distinction is rarely made. In his inventory of phonatory settings, Laver (1980: 122-126) reviews a number of key properties of harsh creaky voice. The acoustic properties of harsh creaky voice include highly irregular cycle-to-cycle variations in the duration of the pitch periods, dicrotic and tricrotic phonation (period doubling and tripling), and low fundamental frequencies usually below 100 Hz. The motoric properties of harsh creaky voice involve thickened and somewhat compressed vocal folds, typically with recruitment of the ventricular folds. The vocal folds are often slack while the values of subglottal pressure and airflow are lower than those of modal voice (see Blomgren et al., 1998). This indicates that the production of harsh creaky voice mostly likely depends on the coordinated action of mechanisms (2) and (3), but not (4).

The phonation type of harsh voice is likewise reviewed by Laver (1980: 126-132). The acoustic properties of harsh voice include highly irregular period-to-period variations in duration (jitter) and amplitude (shimmer), greater intensity, and fundamental frequencies typically above 100 Hz. The motoric properties of harsh voice comprise
tightly compressed vocal folds often accompanied by hyperadducted ventricular folds. Characteristically the vocal folds are tense while the driving pressure is higher than for pure voice. Harsh voice is lexically distinctive on vowels in the Tibeto-Burman language Bai, together with tone and nasalization. Edmondson et al. (2000) made video recordings of near minimal vowel contrasts with a stroboscopic laryngoscope. The laryngeal images of harsh voice demonstrate that not only was there participation of the ventricular folds, but also the aryepiglottic folds situated above them are constricted in a sphincter-like manner (see Figure 3.2 for the relative position of the aryepiglottic folds).\(^3\) Since high subglottal pressures are needed to overcome the flow resistance of a tense, constricted, and longitudinally extended laryngeal pathway, harsh voice appears to depend on the combined effect of mechanisms (2) and (4), as suggested by the Jiang et al. 2001 study. In addition, at least some degree of spatial nonuniformity (3) might be expected when the vocal, ventricular, and aryepiglottic folds all play a role in the production of harsh voice.

**Figure 3.2** Anatomical positions of the aryepiglottic folds, ventricular folds, and vocal folds viewed posteriorly.
As the aryepiglottic folds are currently not included in computational models of the vocal folds, their effects on the vibratory pattern of harsh voice must await future research.

The preceding account shows that the phonation types of harsh and harsh creaky voice share the common properties of highly irregular pitch periods and small glottal rest area. It appears then that modulation noise can serve as an acoustic cue for tight adduction of the vocal folds. Modulation noise can therefore be defined as the perturbations of the glottal wave resulting from the nonlinear behavior of the vocal folds in a tightly adducted state.

3.2.2 Measures of modulation noise

Over the past five decades diverse methods have been developed to estimate modulation noise (Buder, 2000), though there has been one fairly recent attempt to provide a framework that unifies perturbation measures (Pinto and Titze, 1990). The most used and widely investigated measurements are those assessing the unpredictable frequency (jitter) and amplitude (shimmer) modulation of the glottal waveform. Typical measures of frequency and amplitude modulation noise are respectively:

\[
\text{Percent jitter (\%)} = 100 \times \frac{1}{N - 1} \sum_{i=1}^{N-1} |T0_i - T0_{i+1}| \bigg/ \frac{1}{N} \sum_{i=1}^{N} T0_i,
\]

\[
\text{Absolute shimmer (dB)} = \frac{1}{N - 1} \sum_{i=1}^{N-1} |20 \log_{10} \left( \frac{A0_{i+1}}{A0_i} \right)|,
\]

where \(T0\) and \(A0\) are the fundamental period and amplitude of the individual cycles.

Hollien et al. (1973) showed that percent jitter ranged from 0.33 to 1.34\% in normal speakers, the mean jitter being 0.97\%. There appeared to be a slight tendency for
the percent jitter to increase with rising fundamental frequency, a finding subsequently confirmed by Horii (1979). In a later investigation, Horii (1980) found that the means for percent jitter and absolute shimmer were 0.64 % and 0.39 dB, with some vowel quality interaction. However, Horii (1982) examined eight different vowel qualities in a follow-up throat accelerometer study and observed no significant between-vowel differences in jitter or shimmer. Stochastic variations in airstream and vocal fold parameters produce small perturbations of the pitch period not subject to direct linguistic control (Schoentgen, 2001). As the above results indicate, the probable upper limit of these fluctuations is the jitter ($\leq 1\%$) and shimmer ($\leq 0.5$ dB) of normal voice. On the other hand, large perturbations occur during the production of creaky voice. Horii (1985) reported that in healthy speakers the mean jitter and shimmer values of creaky voice (2.47%, 1.15 dB) differed significantly from those of normal voice (0.87%, 0.48 dB).

Although the jitter and shimmer values of creaky voice are rather large, comparable amounts of measured jitter and shimmer can be produced by moderately high levels of additive noise. For example, Hillenbrand (1987) synthesized an unmodulated 130 Hz voice signal with a HNR of 8 dB. Despite the fact there was no actual jitter or shimmer, the measured jitter and shimmer of the signal was 2.61% and 1.17 dB, remarkably close to the means observed for creaky voice. There are two further difficulties with jitter and shimmer measures. First, they are only weakly independent, with Hillenbrand’s study showing that synthesized jitter has more of an effect on measured shimmer than vice versa. Second, they require 20 to 30 pitch periods to obtain a stable mean (Titze et al., 1987), a severe disadvantage in connected speech.
Using an inverse filtering technique to obtain the voice source (residue) signal, Davis (1976) proposed six acoustic measurements for detecting laryngeal pathology. They are the pitch (PPQ) and amplitude (APQ) perturbation quotients, the spectral flatness of the residue signal (SFR) and the inverse filter (SFF), the coefficient of excess (EX), and the pitch amplitude (PA). The PPQ and APQ consist of moving average estimates of the jitter and shimmer of the residue. Spectral flatness is the logarithm of the ratio of the geometric mean to the arithmetic mean of the spectrum; its magnitude increases with rising noise levels (Gray and Markel, 1974). As a result, the SFR and SRF indicate the masking of the harmonic and formant structures by noise, respectively. The EX gives the approximate signal-to-noise ratio of the residue. The PA is the value of the inverse filtered autocorrelation coefficient peak corresponding to the fundamental period. The PA therefore decreases whenever the waveform deviates from perfect periodicity at the correlation delay associated with the fundamental frequency. Consequently, one would expect its value to decline with 1) increasing jitter and shimmer and 2) increasing additive noise. From signal processing considerations alone, therefore, it is clear that the PA is orthogonal neither to the perturbation quotients nor to the SFR. The correlation matrices for normal and pathological subjects bore out this premise. There were highly significant negative correlations between the PA and the SFR of the normal speakers and between the PA and the SFR, PPQ, APQ of the pathological speakers. In spite of the lack of independence of the six acoustic measures, fairly good separation between normal and pathological voices was achieved using a maximum likelihood classifier.

In a later study of pathological voice employing the same six measures, Prosek et al. (1987) found similarly high negative correlations between the PA and the SFR, PPQ,
A perceptual test of voice quality was also carried out, the results of which were correlated with the acoustic findings. Of all the residue measures, the PA performed the best, yielding the highest correlations with scales representing breathy, hoarse, harsh, and strangled voice qualities. Eskenazi et al. (1990) used multiple linear regression to examine the relationship between several acoustic parameters and listener-rated pathological voices. The acoustic parameters included the PA, percent jitter, and a time-domain harmonics-to-noise ratio (HNR). The regression analysis indicated that the PA and the time-domain HNR were the main factors associated with vocal fry; percent jitter was found to be the best predictor of breathiness. Parsa and Jamieson (2000) conducted an investigation comparing the effectiveness of residue and HNR measures as discriminators between normal and pathological voices. The SFR and PA presented significantly better mean classification rates (96.5 and 93%) than the HNR measures (≤ 83.3%).

The preceding studies demonstrate that the PA is a very sensitive indicator of waveform fluctuations, whether they arise from modulation noise or from the additive noise of breathiness. This sensitivity suggests that a peak autocorrelation measure like the PA would be an excellent means of quantifying modulation noise if its response to additive noise could be severely curtailed. Three additional characteristics of the peak autocorrelation coefficient (pACC) make it attractive as a possible cue for modulation noise:

1) It has been widely used in voicing and pitch determination algorithms, either alone or in conjunction with other acoustic parameters (Hess, 1983).
2) The energy normalization procedure has a straightforward physical interpretation when compared to those used in cepstral analysis.

3) The parameter is consistent with the autocorrelation model of auditory pitch processing.

### 3.2.3 Mechanisms of aspiration noise

Aspiration noise is generally considered to arise from a broadband acoustic source produced by turbulent aerodynamic flow in the vicinity of glottis, whereas frication noise is associated with a constriction in the supralaryngeal vocal tract (for a review of the various definitions of aspiration noise, see Jackson, 2000: 9-10). Laminar airflow occurs when a particle stays on the same streamline and has the same velocity as the other particles in the flow. In turbulent airflow, on the other hand, the fluid motion displays unsteady or random behavior. An important parameter for turbulent flow is the dimensionless Reynolds number

\[ \text{Re} = \frac{Vd}{\nu}, \tag{3.5} \]

where \( V \) is the flow velocity, \( d \) the characteristic dimension, typically the diameter of a tube, and \( \nu \) the kinematic viscosity of air, about 0.15 cm\(^2\)/s. When the Reynolds number exceeds a certain critical value, on the order of 1700 ± 200 for speech, the laminar flow becomes turbulent (Catford, 1977). The Reynolds number is the ratio of the inertial force (mass \( \times \) acceleration) to the viscous damping force. At low Reynolds numbers, small perturbations in the flow are quickly damped out downstream. However, as the Reynolds number approaches the critical value, any spatial disturbance may lead to a laminar-to-turbulent flow transition.
Acoustic noise sources are generated by aerodynamic turbulence. Because the relationship between acoustic noise sources and aerodynamic turbulence is highly complex (Blake, 1986), the realistic modeling of noise sources along the length of the vocal tract presents serious challenges. However, as will be shown in the next chapter, a simplified model of aspiration noise appears to be adequate for the purposes of estimating the harmonic-to-noise ratio. To develop such a model, much of the discussion that follows concerns the shape and falloff rate of the spectrum of aspiration noise.

Acoustic sources created by turbulent airflow have been represented as consisting of varying proportions of three components: a monopole, a dipole and a quadrupole source (Goldstein, 1976). A monopole noise source is a vanishingly small pulsating sphere that radiates acoustic waves equally in all directions. A dipole source may be understood as two monopole sources pulsating in opposite phase and separated by a small distance relative to the wavelength. A quadrupole source is made up of two opposite dipoles. Stevens (1998: 100-101) offered an assessment of how each of these sources contributes to turbulent noise generation in the vocal tract (see also Shadle, 1985: 21-23).

1) The velocity fluctuations resulting from abrupt spatial changes along the length of the exit tube (glottis) constitute a monopole source.

2) The fluctuating forces due to the jet of air impinging on a rigid obstacle or hard surface (ventricular folds and epiglottis) form a dipole source.

3) The velocity fluctuations generated by the free jet of air in the space downstream from the flow exit plane (glottis) represent a quadrupole source.
In comparison to the monopole and dipole sources, the quadrupole appears to be a relatively inefficient source of turbulence noise in speech and has not been modeled in electrical analogs of the vocal tract (Shadle, 1985: 22; Stevens, 1998: 101).

Because of the complex geometry and relative inaccessibility of the vocal tract, mechanical tube models have been used to investigate the spectral characteristics of noise sources. Pastel (1987) mounted a uniform circular nozzle of area 0.114 cm\(^2\) at the closed end of a circular tube 17 cm long and 2.85 cm\(^2\) in cross-sectional area. The nozzle was positioned at right angles to the long axis of the tube. The volume velocity was fixed at 442 cm\(^3\)/s.\(^6\) The noise source spectra were obtained by measuring the sound pressure 26 cm downstream from the mouth of the tube and subtracting the calculated tube transfer function. It was found that the source spectrum of the monopole declines slowly up to 1 kHz then rolls off rather rapidly at frequencies above 2 kHz.

Zhang et al. (2002) carried out a modeling experiment that did not require a correction for the tube transfer function. Reflectionless terminations were fitted to the ends of the inlet and outlet sections of a rectangular tube (2.5×2.5 cm\(^2\)) to prevent the formation of standing waves below 7 kHz. A microphone was placed within the tube 14 cm downstream from an orifice plate dividing the two sections. The circular orifice of area 0.114 cm\(^2\) centered in the plate was designed with a converging inlet profile and a sharp-edged outlet. With normal speech pressures from 6 to 12 cm H\(_2\)O, the noise slopes went from roughly –8 dB/octave to –5 dB/octave over the frequency range of 1-7 kHz. The noise slopes produced by the orifice areas of 0.114 cm\(^2\) and 0.317 cm\(^2\) were also compared at 10 cm H\(_2\)O. Both slopes were very similar, but the orifice with the larger cross-section showed a boost in the overall noise level. Note that the mid-band and wide-
band slopes of these noise sources are notably flatter than in Pastel’s experimental setup, even at the lowest static pressure tested (6 cm H₂O).

A monopole noise generator at the glottis may be considered a constant volume velocity source, which couples most effectively to the standing wave resonances of the vocal tract at points of maximum sound pressure. If the vocal folds are sufficiently close together (high glottal shunt impedance), the sound pressure at the glottis is at a maximum for all vocal tract resonances. For each of these resonant modes the monopole is analogical to the constant current source of a RLC parallel circuit with a peak frequency response at the maximum input impedance. Periodic excitation of the vocal tract has likewise been modeled as a constant volume velocity source at the glottis (Fant, 1960: 267).

When a jet of air impinges an obstacle or a surface, the resultant fluctuating forces can be modeled as a dipole source. The intensity of the sound produced depends on the angle at which the airflow strikes the obstacle. If the rigid obstacle is perpendicular to the direction of the airflow, considerable aerodynamic forces are developed in the region where the jet encounters the obstruction. On the other hand, when the obstacle forms a surface more nearly parallel with the direction of the flow, weaker shear forces are distributed along the path of the jet. Stevens (1998: 428-429) posits that such shear forces occur when the glottal airflow impinges on the ventricular folds and epiglottis.

Following theoretical and experimental work on spoilers in high speed flow ducts (Gordon 1968, 1969), Stevens (1971) characterized the source spectrum of the dipole by a very broad maximum two to three octaves wide with a peak frequency (Hz)

\[ f_p = 0.2V / D \], where \( V \) is the flow velocity and \( D \) the width of the obstacle. When the
spectrum level (dB) of the dipole is plotted against the Strouhal number $fD/V$ (the normalized frequency), the curve rises slowly to the peak at 0.2 then drops somewhat more rapidly. For volume velocities and constriction sizes typical of speech, Stevens estimated $f_p$ to range from 500 to 3000 Hz, the lower frequency values being more likely to occur with aspiration noise than with frication noise.

To determine the shape of the dipole source spectrum under speech-like conditions, Shadle (1985: 109-110) constructed a mechanical tube model with an adjustable obstacle. A circular orifice plate containing a constriction with an area of 0.08 cm$^2$ was fitted into the mouth of a round tube 17 cm long and 5.07 cm$^2$ in cross-sectional area. The constriction area of 0.08 cm$^2$ is smaller than the glottal areas observed for aspiration noise (0.1-0.5 cm$^2$, Stevens, 1971) and is more typical of close fricatives. A semicircular obstacle half the area of the tube was attached 3 cm downstream from the constriction. As in Pastel’s monopole experiment, a microphone was installed 26 cm from the mouth of the tube and the source spectrum of the dipole was recovered through inverse filtering. At the volume velocity rates 160, 250, 360, 420 cm$^3$/s, the source spectra exhibit a lowpass characteristic that shifts upward as the flow rate increases. The best regression curve over 0.5-10.0 kHz was found to be a falling exponential of the form

$$p_s = ae^{-bf},$$

where $p_s$ is the pressure source level, $f$ frequency, $a$ a positive constant, and $b$ a negative constant. There was an important difference in intensity between the pure dipole source with the obstacle and the presumed monopole source without the obstacle. When the obstacle was removed, the intensity fell to such a low level that the microphone had to be placed within 6 cm of the constriction for the generated sound to exceed the background noise. The monopole is therefore an inefficient generator of turbulence noise.
compared to the dipole source. The dipole with a full obstacle is estimated to have a noise level 15 dB greater than the monopole source at low frequencies (Stevens, 1998: 107).

In electrical network analogs of the vocal tract the acoustic dipole has been modeled as a noise pressure source introduced in series between two elementary tubes (Fant, 1960: 36; Flanagan, 1972: 54). Just as the volume velocity source couples most effectively to standing wave resonances of the vocal tract at points of maximum sound pressure, a pressure source couples most effectively to the standing wave resonances of the vocal tract at points of maximum volume velocity. The dipole at those points is analogical to a constant voltage source of a RLC series circuit with a peak frequency response at the maximum input admittance. As the dipole is shifted away from the point of maximum volume velocity, the excitation of the mode under consideration decreases proportionately (Stevens, 1998: 167-168).

The dipole sources at the ventricular folds and epiglottis present a variable degree of coupling to vocal tract resonances, depending on the mode and the source location. For instance, a dipole source immediately above the glottis will only weakly excite the two lowest modes (1st and 2nd formants) because the volume velocity at that location is close to zero for those resonance frequencies. This is in contrast with the maximum coupling to all the resonant modes afforded by the monopole source at the glottis. Hence the monopole source may contribute substantially to the overall aspiration noise despite a less efficient sound generating mechanism. On the basis of Shadle’s falling exponential dipole function, Stevens (1998: 430-432) modeled the shear forces at the surface of the ventricular folds and epiglottis as three dipole sources located at points 0.5, 1.5, and 2.5 cm above the glottis, taking into account the coupling of each source to the vocal tract
resonances. His calculations indicate that the dipole sources on the epiglottis (1.5 and 2.5 cm above the glottis) are the main components of the noise spectrum between 3 and 5 kHz, while the dipole source on the ventricular folds (0.5 cm above the glottis) has little spectral effect. The contribution of the monopole source was also computed using Pastel’s source function. The monopole noise source is found to predominate at the lower end of the spectrum, but only below about 1.5 kHz, as Pastel’s function falls off quickly above 2 kHz. It is of interest to point out that the study by Zhang et al. (2002) offers another possible monopole source function that declines much more slowly up to 7 kHz. If the latter source function were adopted, the higher frequencies of aspiration noise would be less dependent on the position and shape of the epiglottis. Further experimental work is obviously needed to clarify this question.

In the source-filter model of the vocal tract, the sound pressure a certain distance from the lips is represented as the product of the mouth volume velocity and the radiation impedance (Fant, 1960). In speech synthesis a simplified radiation characteristic is implemented by applying a first difference filter to the mouth volume velocity, which yields a rise of 6 dB/octave (Klatt, 1980). After removing the all-pole component of the vocal tract transfer function, Stevens (1998: 431) multiplied the combined spectrum of the monopole and dipole sources by the radiation characteristic. The overall spectrum of [h] is fairly flat from 1 to 5 kHz, with a somewhat steeper skirt below 1 kHz than above 5 kHz.

Hillman et al. (1983) investigated the spectral shape of aspiration noise in a more direct manner through a reflectionless tube technique. Subjects covered the mouthpiece of a tube with their lips and produced a neutral whispered vowel. The other end of the
tube was provided with a reflectionless termination so that the spectral properties of the
glottal source could be estimated without the action of the vocal tract filter. The sound
pressure spectra were sampled from 100 to 4000 Hz in steps of 100 Hz. Least-squares
lines were fitted to the spectra, and the resulting slopes were expressed in dB/kHz. A
linear slope measure was employed instead of the more usual logarithmic one in
dB/octave because it showed better correlation values. The overall mean slope for men
and women is –9.4 dB/kHz with a standard deviation of 2.04. No significant differences
were found between the sexes. To determine the effect of the radiation characteristic on
the source spectrum, a typical example of a measured spectrum was given an emphasis of
+6 dB/octave. The modified spectrum was found to be relatively flat over the range of the
first three formants. In light of this result, the authors consider that simple broadband
white noise suffices as an approximation to aspiration noise in synthetic speech. This
view finds further support in Stevens’ model of aspiration noise discussed earlier. In
addition, the speech synthesis algorithm of Klatt (1980) assumes the source spectra of
aspiration and frication noise to be approximately flat, modeling them both as –6
dB/octave broadband noise, the slope of which is effectively canceled by the +6
dB/octave radiation characteristic.

During voiced sounds the vibration of the vocal folds produces an airflow that
varies quasi-periodically. If the glottal opening is small, the periodic component is much
more prominent than the additive turbulence noise so that the harmonic structure is
clearly discernible over the entire frequency range. As the glottis widens and the rate of
airflow increases, the noise gradually swamps the higher harmonics, which are weaker
due to the downward slope of the harmonic envelope. Stevens (1998: 116) calculated that
aspiration noise may rise more than 15 dB above the level of the higher harmonics. The following section reviews a number of algorithms designed to quantify the amount of aspiration noise during voice production.

3.2.4 Measures of aspiration noise

There are two approaches to estimating the relative aspiration noise of speech: frequency domain methods and time domain methods. Frequency domain methods generally estimate the harmonics-to-noise ratio (HNR) by Fourier transforming the speech signal in order to find the ratio between the energy of the harmonic peaks and the energy of the interharmonic valleys. Time domain methods either 1) extract the periodic and noise components directly from the speech waveform or 2) perform a short-term cepstral or correlational analysis to estimate the strength of the periodic component. The latter measurement techniques are based on the same general principles as the PA, the pitch amplitude or inverse filtered peak autocorrelation coefficient, discussed earlier in the context of modulation noise. Note, however, that several algorithms incorporate time domain elements into frequency domain methods and vice versa.

Both time and frequency domain estimation methods are subject to three kinds of error:

1) Poor pitch determination
2) Spectral smearing resulting from the windowing of the speech signal
3) The influence of modulation noise (jitter and shimmer) on measures of aspiration noise
The first two errors result in the loss of power of the original harmonic components to other frequencies (spectral leakage). The last error is the most serious because it can obscure linguistically significant contrasts between phonation types with considerable aspiration noise (breathy voice), considerable modulation noise (harsh modal or harsh creaky voice), or both (harsh whispery voice). Measures of modulation noise face an analogous problem due to the confounding presence of aspiration noise (see Section 3.2.2). Consequently, the following account will consider not only the basic analysis techniques, but also performance in the presence of modulation noise.

Frequency domain methods use the discrete Fourier transform to measure the HNR. For example, Kojima et al. (1980) selected three consecutive pitch periods and analyzed the sequence by means of a Fourier series expansion. The squared magnitudes of the harmonics and the squared magnitudes of the noise components between the harmonics were calculated, then the ratio of their respective sums was formed. Pitch and windowing errors were kept to a minimum by manually selecting the three-period sequence, but at the cost of slow computation time. The HNR of their normal subjects ranged from 15.0 dB to 23.5 dB, while the mode of the distribution was 20 dB. No significant difference was found between men and women. Kasuya et al. (1986) applied a somewhat similar technique to obtain the normalized noise energy (NNE), an inverse automated measure of the HNR. The NNE values of normal subjects ranged from about –6 dB to –26 dB, with the mode approaching –15 dB. Klingholz (1987) developed an analysis-by-synthesis variant of the frequency domain method. A spectral fitting technique determined the peak frequencies and bandwidths of the harmonics up to 5 kHz. These parameters are used to construct a synthetic harmonic spectrum, the energy of
which is compared to the total energy of the Gaussian windowed speech. The HNR range of the normal voices was from 15 dB to 24 dB, a result nearly identical to that of the Kojima et al. study; normal men and women exhibited a mode at about 24 dB. These studies indicate that the upper bound of the HNR is on the order of 24 dB for normal voiced speech. A number of researchers (Muta et al., 1988; Cox et al., 1989) have noted that frequency-domain HNR estimates based on the Kojima et al. method are degraded by modulation noise. The explanation for this phenomenon lies in the fact that jitter and shimmer give rise to additional frequency components around each harmonic, the former having a substantially greater effect than the latter (Klingholz and Martin, 1985; Hillenbrand, 1987, Murphy, 2000). Using a maximum likelihood approach, Jackson and Shadle (2001) decomposed the speech signal with a pitch-scaled harmonic filter to obtain estimates of the HNR. Tests with synthesized signals showed that large amounts of jitter ($\geq 1.5\%$) and shimmer ($\geq 1.5$ dB) had a lowering effect on measured HNRs, particularly when the actual HNR was above 20 dB.

The basic time domain algorithm to compute the harmonics-to-noise ratio was developed by Yumoto et al. (1982). The algorithm cancels the noise component by averaging a sequence of pitch periods, thereby yielding a waveform that represents the harmonic component. The variances of the harmonic waveform relative to each pitch period are then averaged to provide an estimate of the noise energy. The mean HNR was found to be 12.2 dB and 11.5 dB for normal men and women with a range of 7.0-17.0 dB. Awan and Frenkel (1994) utilized a very similar method and obtained an overall normal mean of 15.4 dB with a range of 13.6-18.2 dB. Note that the upper HNR bounds for normal speakers are appreciably lower than those reported for the frequency domain.
estimation techniques. These results probably reflect the greater sensitivity of time
domain algorithms to modulation noise than is the case for frequency domain algorithms.
The basic time-domain method has been modified in several ways. Ladefoged and
Antoñanzas-Barroso (1985) used only part of each pitch cycle in the sequence, thereby
minimizing the potential bias from different sized periods. Qi (1992) applied dynamic
programming to align the noise components in time before calculating the HNR.
Evaluating the method with a synthetic vowel, he found that the HNR remained relatively
constant (± 2 dB) with increasing jitter. Yet he cautioned that all variations due to
temporal differences may be eliminated, including those resulting from additive noise.
Murphy (1999) proposed an extensive modification of the Yumoto et al. technique
whereby waveform samples are replaced with pitch synchronous Fourier coefficients. As
the spectra of the individual cycles are added by harmonic number and not by frequency
location, period deviations are effectively removed. Tests with synthetic voice signals
showed that the normalized HNR was insensitive to shimmer and only slightly degraded
by jitter.

Other time domain approaches to estimating the strength of the periodic
component include cepstral and correlational techniques. The cepstrum is obtained by
calculating the Fourier transform of the time waveform, taking the logarithm of the
magnitude of the transform, then performing an inverse Fourier transform on the
logarithmic function. The autocorrelation of a signal can be computed either directly in
the time domain or, like the cepstrum, by means of a dual Fourier transform. As the
degree of periodicity of the voice source grows, both the cepstral and correlation
functions display increasingly higher peaks at time delays equal to the fundamental
period. Hillenbrand et al. (1994) tested the effectiveness of several acoustic measures of breathiness using a subjective rating scale (direct magnitude estimation) as the basis for comparison. The cepstral peak prominence (CCP) was defined as the maximum cepstral peak normalized by a regression line technique. There were very high negative correlations between the breathiness rating and each of his three cepstral measures. Blankenship (1997) corroborated these findings by showing that the CCP, normalized relative to the mean cepstrum, completely separated breathy from normal voice in Jalapa Mazatec. On the other hand, the CCP was less effective in separating laryngealized from normal voice. A more elaborate cepstral method was developed by de Krom (1993). The cepstral peaks representing the laryngeal source are zeroed by a comb liftering operation, then the resulting cepstrum is Fourier transformed. Because the liftering process removes nearly all the harmonic energy, only the approximate noise spectrum remains. Subtracting the corrected noise spectrum from the original spectrum yields an estimate of relative aspiration noise (see also Qi and Hillman, 1997). Another correlational measure of aspiration noise was implemented by Michaelis et al. (1997). They devised an interband correlation technique, the GNE (glottal-to-noise excitation ratio), which determined the maximum of cross correlations between the Hilbert envelopes of different inverse filtered frequency channels. When the bandwidth of the channels was sufficiently large, the GNE varied only with additive noise and was independent of jitter and shimmer.
3.3 Spectral Tilt

3.3.1 Introduction

Frequently employed measures for assessing spectral tilt are the narrow-band slope $H_1 - H_2$ and the mid-band slopes $H_1 - A_1$, $H_1 - A_2$, and $H_1 - A_3$. Ladefoged and Antoñanzas-Barroso (1985) studied the effectiveness of spectral tilt and the HNR as cues for phonological breathiness of vowels in !Xóõ. The spectral tilt measures were $H_1 - H_2$ and $H_1 - A_1$ while the HNR was estimated using a modified time domain method. The maximum possible separation that could be achieved between pure and breathy vowels depended on the measurement technique: $H_1 - H_2$ wrongly categorized one vowel, $H_1 - A_1$ two vowels, and the HNR three. Thus spectral tilt was found to be a more important indicator of phonological breathiness than aspiration noise. The linguistic relevance of spectral tilt has also been demonstrated for the normal and breathy nasal consonants of Tsonga (Traill and Jackson, 1987). Three measures of spectral tilt were used in this investigation: $H_1 - H_2$, $H_1 - A_2$, and $H_1 - A_3$, the last being defined as the difference between the first harmonic and the most intense harmonic above 2 kHz. The HNR was computed by means of the Yumoto et al. method. A discriminant analysis was carried out in which the probability of successful classification of subjects into groups was calculated. Male and female speakers displayed systematic group variation in how the breathy nasals were realized. Overall, the variables that best modeled the male speakers were $H_1 - H_2$ and $H_1 - A_2$. For female speakers, the optimal variables were $H_1 - A_2$ and HNR. This suggests that in Tsonga aspiration noise may not be the primary cue for male linguistic breathiness.
The various measures of spectral tilt do not always agree. Hanson (1997) found that the narrow-band slope $H_1–H_2$ was not well correlated with the mid-band slopes $H_1–A_1$ and $H_1–A_3$ (see also Hanson and Chuang, 1999). The latter were, however, somewhat better correlated between themselves. The relative independence of narrow-band and mid-band spectral tilts has been attributed to different laryngeal mechanisms. When the phonation type progresses from breathy to laryngealized voice, three changes in the shape of the glottal pulse typically occur:

1) The open quotient diminishes (OQ: the ratio of the open duration to the duration of the period).

2) The pulse becomes less symmetrical.

3) The pulse falls more abruptly (Childers and Lee, 1991; Stevens, 1977; Gauffin and Sundberg, 1989).

Narrow-band spectral tilt appears to be largely dependent on the OQ and the symmetry of the glottal pulse. Mid-band and wide-band spectral tilt, on the other hand, are primarily an indicator of the steepness of the pulse edges.

### 3.3.2 Narrow-band spectral tilt: $H_1–H_2$

Bickley (1982) compared the acoustic spectra of normal and breathy /a/ in !Xóõ and Gujarati. Except for one case in !Xóõ, $H_1–H_2$ successfully kept the phonation types apart. Stevens and Hanson (1995) noted that when there is complete closure during part of the glottal cycle, only the low frequencies are affected by a change in the open quotient. Varying the OQ from 0.3 to 0.7, they found that $H_1–H_2$ increased by about 10 dB if the other parameters of the KLSYN88 speech synthesizer remained constant (Klatt...
The finding was confirmed by Holmberg et al. (1995) who measured the closed quotient (1–OQ) as well as H1–H2, H1–A1, and H1–A3. Twenty women recorded tokens of the vowel /æ/, and their aerodynamic data and acoustic spectra were compared. There was a close negative correlation between H1–H2 and the closed quotient; however the correlations were notably weaker between the closed quotient and the mid-band slopes H1–A1 and H1–A3. Sundberg et al. (1999) found an even stronger negative correlation between H1–H2 and the closed quotient of the glottal airflow when they examined the syllables [pa] and [pæ] sung by five baritones.

The monotonic relation between H1–H2 and OQ may be confounded by the spectral consequences of other glottal flow parameters. Fant et al. (1985) introduced a popular four parameter model of the time differentiated volume velocity pulse $U_g'(t)$, known as the LF model (see Figure 3.3 above). The first segment of the model can be
most simply described as three-quarters of a sine wave multiplied by a negatively damped exponential factor. The second segment is an exponential curve joined piecewise at the point of discontinuity then attenuating to zero. One of the parameters, the symmetry ratio \( R_k \), is the relative duration of the falling branch from the peak \( (T_p) \) to the point of discontinuity \( (T_e) \). Like the OQ, an increase in \( R_k \) augments the amplitude of the fundamental (Fant, 1995). After deriving analytically the frequency domain expression for the LF model, Doval and d’Alessandro (1997) showed that the same value of \( H_1–H_2 \) could be achieved by jointly varying OQ and \( R_k \) over their respective ranges (see also Henrich et al., 2001). There is also direct experimental evidence for the effects of OQ and \( R_k \) on \( H_1–H_2 \). Swerts and Veldhuis (2001) used inverse filtering to determine the OQ and \( R_k \) of the Dutch vowel /a/. Of the 28 observed cases, there were 25 positive correlations between \( H_1–H_2 \) and \( R_k \), but only 17 between \( H_1–H_2 \) and OQ. On the other hand, \( H_1–H_2 \) and the closed quotient of glottal flow significantly differentiated the lexically contrastive breathy and nonbreathy tones of Hmong, whereas the waveform symmetry measure did not (Huffman, 1987).

### 3.3.3 Mid-band and wide-band spectral tilt

During normal voice the glottal flow spectrum is considered to fall off at the asymptotic rate of about –12 dB/octave (Flanagan, 1957). This slope, for example, characterizes the glottal pulse functions used in the KLSYN and KLSYN88 speech synthesizers (Klatt, 1980; Klatt and Klatt, 1990). If the radiation impedance is assumed to be approximately proportional to frequency, then the radiated sound pressure is related to the mouth volume velocity by a 6 dB per octave rise in the frequency domain and a
differentiation process in the time domain. Thus for normal voice, the observed wide-band spectral slope should be approximately \(-6\) dB/octave if the effects of the vocal tract transfer function are neglected. As discussed earlier, Fant et al. (1985) developed a model of differentiated glottal flow that included the OQ and the symmetry parameter \(R_k\).

Another parameter, the return time constant \(T_a\), is defined as the temporal difference between the point of discontinuity and the projection on the time axis of the slope of the exponential curve (Figure 3.3). By varying \(T_a\), one modifies the duration of the closure phase of glottal flow. If \(T_a\) is set to 0, then the wide-band slope of the differentiated pulse spectrum is the \(-6\) dB per octave typical of normal voice. If \(T_a\) is greater than 0, the wide-band slope decreases to the \(-12\) dB per octave representative of lax or breathy voice, while the cutoff frequency of the additional lowpass attenuation takes on the value of \(1/2\pi T_a\) (Fant 1995). The model shows that the spectral envelope of the higher harmonics is highly sensitive to the abruptness of the closure phase. From an acoustic-motor point of view, when the vibrating vocal folds begin to separate, a triangular chink is created between the arytenoid cartilages. The acoustic mass and resistance of the air within this volume act like a lowpass filter operating on the rapidly varying components of the flow waveform (Stevens and Hanson, 1995; Hanson, 1997). As the folds are pulled farther apart (as in breathy voice), airflow leakage over the length of the glottis causes additional smoothing of waveform discontinuities. Thus the net effect of glottal widening is to reduce the amount of energy in the mid to high harmonics (Cranen and Schroeter, 1995).

There are two phenomena not accounted for in the LF model. First, the high frequency tilt of the differentiated glottal flow may be flatter than the \(-6\) dB/octave of normal voice, in particular during laryngealized phonation. For this reason, the present study models
wide-band slopes from 0 to –18 dB/octave. Second, high speed cinematography has revealed instances where the opening phase of the glottal pulse takes place more rapidly than the closure phase (Hirano et al., 1983). Since the magnitude spectrum of a time reversed pulse is identical to that of the original waveform, the spectral slope should be insensitive to phase relationships, presumably like the ear itself. Hence the spectral tilt has the advantage of providing an indication of the steepness of the glottal pulse branches, but not the temporal order in which they occur.

The mid-band slopes H1–A1, H1–A2, and H1–A3 have been used to estimate spectral tilt. The question arises as to which of the three best distinguishes among the different phonation types. Blankenship (1997) tested a number of measurement techniques in a trial study of Jalapa Mazatec, which contrasts breathy, normal, and laryngealized vowels. Fourteen samples of each phonation type were selected, then the spectral tilts H1–A1, H1–A2, H1–A3 were measured. Both H1–A2 and H1–A3 were able to distinguish laryngealized voice from normal voice in 13 of the 14 cases. However when it came to separating breathy from normal voice, there was a sharp difference: H1–A2 performed correctly in 12 cases, while H1–A3 succeeded in only 7. Furthermore, H1–A1 did not provide an effective cue for laryngealized voice (laryngealized correct: 9/14; breathy correct: 13/14). Thus H1–A2 appears to be the best measure of spectral tilt for this data set. Also in Gujarati H1–A2 yielded fewer errors than H1–A3. For each of the five speakers, the measures H1–H2, H1–A1, H1–A2 kept the normal vowels /a/ and /e/ distinct from their breathy counterparts; H1–A3, however, resulted in two misclassifications (Fischer-Jørgensen, 1967).
Jackson et al. (1985) developed an inverse filtering technique to measure spectral tilt. The method, similar in spirit to the one proposed here, involves fitting a regression line to the amplitudes of the harmonics. Vowel phonation types were investigated in three languages: Hmong /a/ (breathy vs. normal), Jalapa Mazatec /e/ (breathy vs. normal vs. laryngealized), and Burmese /a/ (normal vs. slightly laryngealized). All the speech data were inverse filtered to eliminate the effects of the vocal tract transfer function. After the fundamental frequency was estimated by cepstral analysis, the peak frequencies and magnitudes of the first 8 harmonics were searched. The program then calculated linear regressions on the harmonics H1 through H8 and H2 through H8, both in linear (dB/kHz) and log frequency (dB/octave). Averaged across languages and speakers, the slopes of the three phonation types showed considerable differences according to whether or not H1 was included in the regression model. In the H1-H8 region the log spectral slopes of breathy, normal, and laryngealized phonation were –2.00, –1.87, –1.80 dB/octave respectively, whereas in the H2-H8 region they were –9.14, –7.69, –5.70 dB/octave. Spectral tilt variation in dB/octave was well described by a linear model in both the H1-H8 and H2-H8 ranges. Spectral slope variation in dB/kHz was also well-described by a linear model in the H1-H8 range, but not in the H2-H8 range. However, the differences between the mean slopes of the three phonation types were generally greater in dB/kHz than they were in dB/octave. As the harmonic falloff of the inverse filtered waveform should approximate a logarithmic relationship, the discrepancy was explained by the effects of subglottal resonances and individual differences. Another discrepancy, the lack of agreement between the slopes of normal voice and the normative –12 dB/octave high
frequency tilt of the residue signal, might possibly be due to imperfect inverse filtering, restricting the analysis range to the eighth harmonic, or a combination of both.

Spectral tilt has also been assessed by using wide-band measures. Hammarberg et al. (1986) divided the spectrum into three frequency bands, 0-2 kHz, 2-5 kHz, 5-8 kHz, and found the peak amplitude in each one. Listener evaluations of normal and pathological voices showed that breathiness was associated with a rapid decline in energy from the 0-2 kHz band to the 2-5 kHz band. The level of the 5-8 kHz band was found to be only slightly less than that of the 2-5 kHz band. The relatively high energy in the uppermost band was considered due to the nearly flat spectrum of the turbulent noise source boosted by the +6 dB/octave radiation characteristic.

3.4 Summary and conclusion

This chapter examines the mechanisms underlying modulation and aspiration noise as well as the methods currently used to estimate their relative levels. The glottal characteristics associated with spectral tilt are likewise considered and the relevant acoustic measures discussed. Modulation noise arises from the cycle-to-cycle perturbation of the pitch period resulting from the greater than normal adduction of the nonlinear vocal fold system. Modulation noise characterizes in particular two phonation types: harsh creaky voice and harsh (modal) voice. Harsh creaky voice (or laryngealized creaky voice) is brought about by compressed and rather slack vocal folds, most often with some ventricular participation. Harsh (modal) voice is produced by a tightly adducted laryngeal pathway, involving not only the vocal and ventricular folds, but often the aryepiglottic folds as well. The temporal consequence of modulation noise is the
unpredictable frequency (jitter) and amplitude modulation (shimmer) of the glottal wave; its spectral consequence is the appearance of additional frequency components around each harmonic. Aspiration noise is hypothesized to be generated by distributed turbulent sources, a monopole at the exit of an abducted glottis and dipoles at the surface of the ventricular folds and epiglottis. Assuming the radiation at the lips is taken into account, aspiration noise displays a relatively flat spectrum, the upper region of which tends to swamp the higher harmonics of the quasi-periodic glottal source. For normal speakers, the upper limit of the harmonics-to-noise ratio (HNR) is about 24 dB while the lower limit is on the order of 15 dB. The wide-band slope of normal voice is approximately –6 dB/octave when radiation effects are considered. As the size of the glottal aperture grows larger, the wide-band slope becomes steeper than –6 dB/octave.

There are several questions to be addressed regarding measures of laryngeal noise and spectral tilt. The most general issue that confronts any method that attempts to quantify laryngeal parameters is its plausibility from an auditory neural perspective. A number of acoustic analysis systems have been devised to estimate these parameters, but only one can claim to represent the auditory periphery. In a study of pathological breathiness, Shrivastav (2001) implemented a peripheral auditory stage whose output was used to calculate measures of aspiration noise and spectral tilt. He found that the measures computed with the auditory model were better predictors of perceived breathiness than those without the model. While there are many uncertainties concerning central neural mechanisms, several domains of speech processing (speech recognition, acoustic-to-articulatory inversion) necessitate a lookup procedure to achieve a fast and computationally efficient solution, and it is likely that the brain performs analogously as
well. To determine the relative aspiration noise and wide-band slope, a table lookup technique is proposed: Analysis-by-Synthesis Harmonic estimation (ASHE). The procedure is similar to the regression line method of modeling harmonic slope as it uses least squares spectral matching (cf. the Jackson et al. study in Section 3.3.3). However the method also provides a simultaneous estimate of the HNR. A further advantage of ASHE is that it gives a fully automated running measure of wide-band spectral tilt. By comparison, the mid-band spectral tilts H1–A1, H1–A2, H1–A3 require hand measurement of formant peaks as automatic formant tracking algorithms are prone to error. In the course of this background chapter, the question was raised as to the separability of modulation and turbulence noise. Both types of laryngeal noise may involve spectra with the harmonic peaks surrounded by broadband continuous noise. Because the two types of laryngeal noise are phonologically contrastive, it is important to find acoustic measures particularly sensitive to either modulation noise or aspiration noise, but not both at once. In Chapter 4 a band-limited autocorrelation function designed to achieve this goal is presented and its ability to discriminate between modulation and aspiration noise is tested in Chapter 5.
NOTES TO CHAPTER 3

1. This summary follows Nayeh and Mook (1979), particularly Chapter 4.

2. Discontinuities are also introduced when the excitation frequency $\omega$ is fixed and the excitation amplitude $A$ is varied (Nayfeh and Mook, 1979: 169).

3. The aryepiglottic folds contain the aryepiglottic muscles which pull the epiglottis down over the glottis.

4. To account for the constricted vibration patterns of throat singing, Sakakibara et al. (2001) simulated the vocal and ventricular folds with a 2×2-mass model.

5. Hollien et al. (1973) used the fundamental frequency variable $F_0$ instead of $T_0$ in Equation 3.3. The percent jitter values are approximately the same for both variables (see Horii, 1979).

6. A volume velocity of 442 cm$^3$/s and an orifice area of 0.114 cm$^2$ is equivalent to a pressure drop of 8.74 cm H$_2$O when $P_d = \rho U^2 / 2A^2$; $\rho$ is $1.14 \times 10^{-3}$ gm/cm$^3$, $U$ the volume velocity, $A$ the orifice area, and 1 cm H$_2$O is 980.665 dyne/cm$^2$ (Flanagan, 1972: 55).
CHAPTER 4

SPEECH ANALYSIS METHODS

4.1 Introduction

In this chapter the observations and conclusions presented in the preceding chapter are further developed and expanded. It was pointed out that methods providing a more realistic auditory neural representation are to be preferred over those that do not. To this end, a stage of auditory filtering is implemented. The F0 extraction technique utilized in this study is based on a band-limited autocorrelation function. The peak reciprocal time delay of the autocorrelation function specifies the fundamental frequency (F0) while the logarithm of the normalized peak autocorrelation coefficient (log pACC) is put forward as an indicator of modulation noise. Furthermore, the narrow-band slope H1–H2 is computed automatically using this F0 value. The HNR and wide-band harmonic slope are used to estimate aspiration noise and overall spectral tilt, respectively. The Analysis-by-Synthesis Harmonic Estimation algorithm (ASHE) assesses the HNR and harmonic slope simultaneously by searching the best match between the short-time speech spectrum and F0 dependent model harmonic spectra. Lastly, an auditorily plausible method of approximating Q of the first formant is developed as well as a set of criteria to determine the boundaries of phonetically homogeneous subsegments in the speech wave.
4.2 Pitch and Voicing Determination

4.2.1 Auditory filter bank

The first stage of analysis is a bank of linear digital resonators, each of which has the transfer function

\[
H(z) = \frac{1 - 2e^{-\pi B_{3dB} T} \cos(2\pi f_d T) + e^{-2\pi B_{3dB} T}}{1 - 2e^{-\pi B_{3dB} T} \cos(2\pi f_d T) z^{-1} + e^{-2\pi B_{3dB} T} z^{-2}},
\]

where \( f_d \) is the damped frequency (Hz), \( B_{3dB} \) the 3 dB bandwidth (Hz), \( T \) the sampling interval (s), \( f_r \) the resonant frequency (Hz), and \( Q_{3dB} \) the 3 dB quality factor. The impulse response is identical to that of a second order analog resonator sampled at times \( nT \) (Gold and Rabiner, 1968; Klatt, 1980). To realize this filter, \( B_{3dB} \) must be specified for each \( f_r \).

Auditory filter bandwidths have been measured by means of a notched noise method where a curve function is fitted to signal-to-masker thresholds at varying notch widths. Using this technique, Glasberg and Moore (1990) determined a formula relating the equivalent rectangular bandwidth (\( ERB \)) to the center frequency in kHz (\( F \))

\[
ERB = 24.7(4.37F + 1).
\]

This equation may be rewritten as

\[
ERB = \frac{f_c}{Q_{max_{ERB}}} + B_{min_{ERB}}; Q_{max_{ERB}} = 9.26, B_{min_{ERB}} = 24.7,
\]

where \( f_c \) is the center frequency in Hz, \( Q_{max_{ERB}} \) the asymptotic quality factor at higher frequencies, and \( B_{min_{ERB}} \) the minimum bandwidth. If the resonance curve does not deviate far from symmetry (low damping), the following theoretical relation holds between the \( B_{3dB} \) and \( ERB \).
\[ B_{3dB} = \frac{2}{\pi} ERB \]  

(4.4)

Consequently, one can express \( B_{3dB} \) as a function of \( f_r \)

\[ B_{3dB} = \frac{2}{\pi} \times 24.7(0.00437 f_r +1); \quad Q_{max_{3dB}} \approx 14.55 \]  

(4.5)

where \( Q_{max_{3dB}} \) is the asymptotic quality factor defined at the half power points.

The 10 dB quality factor (\( Q_{10dB} \)), the resonance frequency divided by the bandwidth 10 dB above threshold, is a measure of frequency selectivity often used in psychophysical and physiological tuning curve studies. Again assuming there is no marked asymmetry in the resonance curve, one can establish the approximate relation

\[ Q_{10dB} = Q_{3dB} / 3, \]  

(4.6)

which leads to

\[ Q_{max_{10dB}} = Q_{max_{3dB}} / 3 \approx 4.85. \]  

(4.7)

This value compares favorably with tuning curves widths found in humans as well as in other vertebrates. At 4.0 kHz, human psychophysical tuning curves show mean \( Q_{10dB} \) values of 6.48 for adults and 6.79 for infants; at 2.0 kHz, the \( Q_{10dB} \) values are respectively 5.0 and 5.8 (Olsho, 1985). The auditory nerve fibers of the cat and pigeon display \( Q_{10dB} \) values ranging from 5 to 6 at 4.5 kHz (Sachs et al., 1974). Those of the bobtail lizard have a \( Q_{10dB} \) of about 4 in the vicinity of 4.0 kHz (Manley, 1990). Thus the relationships given in Equation 4.5 may have some applicability across terrestrial vertebrates.

There are indications that the auditory filter may be of higher order than the resonance curve, that is, its top may be flatter and its skirts steeper (Patterson, 1976). To simulate this phenomenon as well as level-dependent nonlinearities, the gammatone and gammachirp models of the auditory filter have been introduced (Irino and Patterson,
The impulse response of a second order resonator is the product of a constant, a damped exponential, and a sinusoid at the carrier frequency. The gammatone function is formed by multiplying this product by the time factor $t^{n-1}$ while the gammachirp function contains both this time factor and an additional frequency modulation term. These functions reduce to the impulse response of a second order linear resonator when $n$ equals 1 and the frequency modulation term is zero. The ERB-based, psychophysical, and physiological quality factors do not deviate markedly from one another when a linear filter model is used to calculate their equivalent asymptotic values. Furthermore, the computational cost of a nonlinear auditory filter is relatively high, especially in view of the generation of the ASHE lookup table discussed below. For the present purposes, then, the linear auditory filter appears to be a good computationally inexpensive approximation. Note that at low frequencies the ERB function increases rapidly, thereby reducing the amplification at resonance ($Q_{3dB}$). This means that the highpass rolloff of the ear is already modeled to some extent, so no additional filtering is implemented.

Creaky voice can occur at frequencies as low as 10.9 Hz, although the typical range is between 20 Hz and 70 Hz with a mean of about 50 Hz (Blomgren et al., 1998). To account for possible subharmonic frequencies, the limits of the filter bank are fixed at 12.69 Hz ($16 \, \text{Hz} / 2^{1/3}$) and 6501.99 Hz, a span of nine octaves. Centered within each octave $l$ are $m$ frequencies spaced at equal intervals ($2^{1/m}$) on a log$_2$ scale. For example, octave 1 ($m = 6$) consists of 6 frequencies 1/6 octave apart, of which the first (13.45 Hz) and last (23.97 Hz) are respectively 1/12 octave ($2^{1/2m}$) above and below the octave endpoints (12.69 Hz, 25.4 Hz). The 330 resonant frequencies ($f_r$) of the filter bank are
given by the following \((l, m)\) pairs: \((1, 6), (2, 12), (3, 24), (4, 48), (5, 48), (6, 48), (7, 48), (8, 48), (9, 48)\). Thus once \(f_r\) exceeds 101.59 Hz \((l \geq 4)\), the number of frequencies per octave (48) and the separation interval \((2^{1/48})\) remain the same. The smallest separation interval of \(2^{1/48}\) (1.45% or one-eighth tone) is on the order of the frequency difference limen for F2 (1.5%, Kewley-Port and Watson, 1994). This value is larger than the fundamental frequency DLs for low pitched click trains (< 1%, Cullen and Long, 1986) and synthetic vowels (0.3-0.5%, Flanagan, 1972: 281), but smaller than the DL for the linguistic significance of pitch, either a quarter tone \(2^{1/24}\) (2.93 %, i.e. song using the 24-step Arabic octave, Touma, 1996) or more probably a semitone \(2^{1/12}\) (5.95%; Hess 1983: 82 proposes a linguistically significant pitch DL of 4-5%).

4.2.2 F0 detection and the peak autocorrelation coefficient (pACC)

Licklider (1951) proposed that the central auditory processor computes the autocorrelation function of incoming neural impulses to create a pitch percept based on temporal interval distributions. Although his specific hypothesis has not been confirmed, the autocorrelation model appears to account for a number of psychophysical and neurophysiological observations (Lyon, 1984; Cariani and Delgutte, 1996a, 1996b; Meddis and O’Mard, 1997; Wiegrebe and Meddis, 2004). In light of this work, a lowpass autocorrelation method is used to estimate the fundamental frequency. The outputs of 188 auditory filters ranging from 13.45 Hz to 830.54 Hz are added together to produce a low frequency summarizing waveform. Current models of the auditory periphery include a nonlinear rectification stage representing the transduction of the filter outputs to neural spikes (cf. Patterson et al., 1995). However, when the output of each of the 188 auditory filters...
filters was subjected to positive half wave rectification before the summation process, the performance of the F0 detection algorithm deteriorated slightly. For this reason, the auditory filter outputs are added linearly. The summed samples are placed inside a 39.37 ms (2^{1/3}/32 Hz) rectangular window shifted every 2.46 ms (1/16 window length). A pitch period 39.37 ms long corresponds to a fundamental frequency of 25.4 Hz. This value lies somewhat below the lower limit of melodic pitch estimated to be about 30 Hz for harmonic complexes up to 800 Hz (Krumbholz et al., 2000; Pressnitzer et al., 2001). The 2.46 ms time step represents a good approximation to the threshold of temporal resolution established by several different methods (Eddins and Green, 1995). After the DC bias is removed from the windowed samples, the short-time autocorrelation function is computed

\[ ACF(k) = \sum_{n=0}^{N-1-k} x(n)x(n+k); \quad ACC(k) = \frac{ACF(k)}{ACF(0)}, \]  

(4.8)

where \( x(n) \) is the windowed speech signal, \( N \) the frame length in samples, and \( ACC(k) \) the autocorrelation coefficient, that is, the autocorrelation function at lag \( k \) normalized by the frame energy \( ACF(0) \). The peak autocorrelation coefficient (pACC) is obtained by scanning the array \( ACC(k) \) by steps of 1 over the lag range \( K_{\text{min}+1} \) to \( K_{\text{max}–1} \). Each \( ACC(k) \) is compared to \( ACC(k–1), ACC(k+1), \) and the previous peak value (initialized to zero at \( K_{\text{min}+1} \)). If \( ACC(k) \) is greater than all three variables, then it replaces the previous peak, otherwise not. The peak picking process continues until it reaches \( K_{\text{max}–1} \), where the last stored \( ACC(k) \) is identified as the pACC. Note that merely finding the maximum value of \( ACC(k) \) causes severe upper harmonic tracking errors when the 39.37 ms window contains fewer than two pitch periods. The largest lag \( K_{\text{max}} \) corresponds to the resonant period of the 26.14 Hz filter. The smallest lag \( K_{\text{min}} \) is set equal to a speaker-
specific F0 designed to reduce the overlap with the first formant: man 285.28 Hz, woman 479.78 Hz, child 570.56 Hz. The upper and lower F0 bounds can also be adjusted from their default values as a means of eliminating gross pitch errors. The reciprocal time lag of the pACC yields an estimate of the fundamental frequency in Hz. As shown in Chapter 5, the algorithm performs relatively well so that no post-processing such as median smoothing is applied to the F0 estimates.

Because there are a finite number of samples in the rectangular window, the amplitude of the pACC falls off as the pitch period increases in length (cf. the upper limit of summation in Equation 4.8; Rabiner and Shafer, 1978: 145). As a result, the fundamental frequencies of the sounds under comparison should be approximately the same if the pACC is to be used as a cue for modulation noise. A number of researchers have proposed cross-correlation methods for pitch analysis (Rabiner and Shafer, 1978: 146-149; Secrest and Doddington, 1983; Medan et al., 1991; Talkin, 1995; Granqvist and Hammarberg, 2003). Like the autocorrelation technique, these methods exhibit a peak at the time lag corresponding to the fundamental period. They furthermore have the advantage that the amplitude of the peak does not decline as the duration of the pitch period increases. Nonetheless, they suffer from the drawback that the normalization process requires the division of the cross-correlation function by a non-constant derived quantity (generally the square root of the product of the shifted and unshifted energies). The term $ACF(0)$ in Equation 4.8, on the other hand, simply represents the energy of the windowed signal. On account of the limited physical interpretability of the normalized cross-correlation, an autocorrelation technique is preferred in the present investigation.
4.2.3 Speech and voice detection

Two parameters play an important role in spoken communication: 1) the presence or absence of speech, and 2) the presence or absence of voice (vibration of the vocal folds). Fant (1960: 230) observed that the speech intensity range is about 30 dB. However, all the recordings used in this investigation showed a range closer to 36 dB.

1) If the frame energy drops below –36.12 dB (relative to the maximum frame energy of the signal) and the frame is not inside a voiceless stop, the frame is labeled [–speech, +silence].

2) If the frame energy drops below –36.12 dB (relative to the maximum frame energy of the signal) and the frame is within a voiceless stop, the frame is labeled [+speech, +silence].

3) Otherwise, the frame is labeled [+speech, –silence].

Recordings with significant levels of background noise would require a narrower threshold range. The frame energy is computed by summing the squared outputs of the 330 auditory filters (12.69-6501.99 Hz) over the 39.37 ms window. In this study, silent regions of the speech signal were often found to contain very low level periodic components. An energy threshold eliminates this type of unvoiced-to-voiced error. Voiceless fricatives likewise produced an occasional false indication of periodicity. As these sounds typically have a high spectral center of gravity, a zero-crossing threshold removes this additional source of unvoiced-to-voiced errors. Accordingly, the voicing decision is determined by several alternative conditions:

1) If the frame is [+silence], the frame is [–voice].

2) If the pACC is less than 0.125, the frame is [–voice].
3) If the pACC is between 0.125 and 0.5 and the frame energy is less than –18.06 dB (relative to the maximum frame energy), the frame is [–voice].

4) If the pACC is between 0.125 and 0.5 and the positive-going zero-crossing rate is greater than 1149.4 Hz, the frame is [–voice].

5) In all other cases, the frame is [+voice].

Note that the energy and zero-crossing thresholds are active only when the pACC lies between 0.125 and 0.5 (–9.03 and –3.01 dB in logarithmic units). The zero-crossing measure is obtained from the linear sum of the 330 auditory filter outputs. The DC bias is subtracted from the sum prior to calculation of the zero-crossings. The threshold values were manually tuned to optimize performance for wide-band speech in low ambient noise conditions. As demonstrated in Chapter 5, the proposed thresholds yield comparatively few voicing errors.

4.3 Narrow-band slope (H1–H2) and estimation of modulation noise

Using the F0 estimate provided by the autocorrelation function, the narrow-band slope H1–H2 is determined automatically. The parameter is defined as the amplitude difference in dB between the calculated values of H1 (the peak frequency between $F_0 \times 2^{-1/12}$ and $F_0 \times 2^{1/12}$) and H2 (the peak frequency between $2F_0 \times 2^{-1/12}$ and $2F_0 \times 2^{1/12}$).

It is important to bear in mind that the narrow-band and wide-band slopes are each a function of the glottal spectrum and the vocal tract filter. As a result, the spectral tilts will unambiguously reflect differences in the glottal spectrum only if the formant structures (and radiation characteristics) are identical. One way to avoid this limitation would be to remove the vocal tract resonances through inverse filtering. However, as
Talkin (1995) emphasized, “when only a few harmonics are present, for instance in voiced obstruents, breathy speech, or falsetto, inverse filtering can remove all traces of the F0 one is trying to estimate.” Thus, errors due to a poor model of the vocal tract filter might overwhelm small differences in spectral tilt.

Modulation noise is estimated by the band-limited pACC (13.45-830.54 Hz) expressed in dB (10log10pACC). It was pointed out above that the finite number of window samples leads to a reduction of the pACC as the pitch period increases in length. Hence the speech sounds to be compared should not only exhibit a similar formant structure, but they should also have roughly similar F0s. It has long been established that a phonetic cue associated with a phonological feature varies as a function of other conditioning acoustic and motoric parameters. Fant (1973: 210) posited that “a minimum requirement of phonetic reality of a feature would be that any two sounds in any context differing by one and the same feature only shall display a difference vector of the same sign along a common phonetic dimension.” Thus the conditioned variability of the modulation noise parameter (log pACC) is not an obstacle to its use as a cue for tight glottal constriction provided that the parameter is compared in the same phonetic context. This statement is equally true for the aspiration noise parameter (HNR) and wide-band harmonic slope since no inverse filter is applied prior to their assessment.

4.4 Wide-band harmonic slope and estimation of aspiration noise

4.4.1 Introduction to Analysis-by-Synthesis Harmonic Estimation

To determine the HNR and the wide-band harmonic slope, the Analysis-by-Synthesis Harmonic Estimation technique (ASHE) finds the minimum Euclidean distance
between the actual speech spectrum and model harmonic spectra over the frequency range of 12.69 to 6501.99 Hz. The lookup table for the model harmonic spectra is generated in three steps:

1) Synthesis of harmonic waveforms
2) Auditory filtering of these waveforms
3) Calculation of the normalized energy of each filter’s output

4.4.2 Synthesis of model harmonic spectra

Harmonic series are constructed by adding together cosine waves whose frequencies are integer multiples of a particular fundamental frequency. There are 156 F0s varying from 26.14 to 570.56 Hz, the values of which are identical to the resonant frequencies of the filter bank. Each F0 has 88 harmonic slopes, the endpoints of which are 0 dB/octave and –18.06 dB/octave. The step size changes according to the range under consideration.

i. Between 0 and –3.01 dB/octave, there are 16 steps of –0.18 dB/octave

ii. Between –3.01 and –6.02 dB/octave, there are 32 steps of –0.09 dB/octave.

iii. Between –6.02 and –9.03 dB/octave, there are 16 steps of –0.18 dB/octave.

iv. Between –9.03 and –18.06 dB/octave, there are 24 steps of –0.37 dB/octave.

The step sizes are intended to provide finer measures in regions most often associated with phonemic contrasts. As was shown in the preceding chapter, flat spectrum noise is a reasonably good model for aspiration noise when the +6/dB emphasis due to the radiation characteristic is taken into account. Wide-band rectangular noise (12.7-6502 Hz) is added to each of the 156×88 harmonic series so as to produce 16 HNR levels. 6

102
The modeled HNRs cover the interval between 0 and 24.08 dB in 1.5 dB increments. The span of 24 dB is chosen in view of the discussion in Section 3.2.4 (cf. Kojima et al., 1980; Klingholtz, 1987). The 156×88×16 harmonic series representing F0, slope, and HNR are synthesized at a 44.1 kHz sampling rate. Each harmonic series is analyzed by 330 auditory filters whose respective outputs are squared and averaged over 2.5 s, thus providing an estimate of the energy in each filter. The 330 energy values are normalized by dividing each of them by their sum, then they are stored successively in a data file. At the end of the synthesis process the file contains 156×88×16 normalized harmonic energy vectors of 330 frequency points (NHEs). In order to speed up the computation, the 156×88×16×330×4 byte binary data file is loaded into memory at the start of the analysis phase.

4.4.3 The HNR and the slope of the short-time speech spectrum

To obtain the HNR and harmonic slope, the algorithm searches for the smallest Euclidean distance between the normalized energy vector of a speech frame and the 88×16 normalized harmonic energy vectors (NHEs) modeling the HNR. Because the F0 of the speech frame is given by the autocorrelation pitch extractor, its value is fixed during the search procedure. The speech waveform is analyzed by the 330 auditory filters whose respective outputs are squared and averaged over the 39.37 ms sliding rectangular window. As in the synthesis stage, the 330 energy values are normalized by dividing each one by their sum, thereby yielding the normalized speech energy vector (NSE). The Euclidean distance between the NSE and the normalized energy vector of any one harmonic series (NHE) is defined as
The minimum Euclidean distance between the NSE and the $88 \times 16$ NHEs is therefore

$$d(NSE, NHE) = \sqrt{\sum_{i=1}^{336} (NSE_i - NHE_i)^2}.$$  \hspace{1cm} (4.9)

The minimum Euclidean distance between the NSE and the $88 \times 16$ NHEs is therefore

$$\min \{d(NSE, NHE_{1,1}), d(NSE, NHE_{1,2}), \ldots, d(NSE, NHE_{n,m}), \ldots, d(NSE, NHE_{88,16})\} \hspace{1cm} (4.10)$$

where the indices of harmonic slope $n$ and HNR $m$ range from 1 to 88 and 1 to 16, respectively. The indices $n, m$ in the minimization function both point to values of the HNR and harmonic slope. As a consequence, once the minimum distance is found, the HNR and harmonic slope are estimated simultaneously.

**4.5 First formant Q and the peak energy factor PE1**

As was stated in Section 2.4.1, the quality factor $Q$ of the first formant ($Q_1$) is equal to the $F_1$ resonance frequency divided by the 3 dB bandwidth ($B_1$). Precise estimation of the bandwidth, and hence the $Q$, of vocal tract resonances has proved a difficult task. Dunn (1961) reviewed a number of problems involved in the direct measurement of bandwidth from the spectral envelope, including the variability of individual judgment, the influence of neighboring formants, and the effects of the harmonic slope of the vowel. He considered the greatest source of uncertainty to lie “in the irregularities of the spectrum of the glottal source.”

For the purpose at hand, only the $Q_1$ of phonation types without vibration of the vocal folds, i.e. glottal noise (glottal stop, whisper, breath), need be approximated. Because the spectral envelope of glottal noise is inherently continuous, the confounding effects of the different harmonic spacings do not exist. Hence some of the uncertainty is removed from the estimation. Nevertheless, this benefit is offset by the stochastic
fluctuations in the glottal noise, although a sufficiently long time window smooths out the more rapid variations of the energy estimate (39.37 ms in this case).

Two methods are currently used to calculate formant bandwidth. The first, implemented in nearly every speech analysis program, relies on linear prediction. Rabiner and Schafer (1978: 450) state:

“A distinct advantage inherent in the linear predictive method of formant analysis is that the formant center frequency and bandwidth can be determined accurately by factoring the predictor polynomial. Since the predictor order $p$ is chosen a priori, the maximum possible number of complex conjugate poles which can be obtained is $p/2$. Thus the labeling problem inherent in deciding which poles correspond to which formants is less complicated for the LPC method since there are generally fewer poles to choose from than for comparable methods of obtaining the spectrum such as cepstral smoothing. Finally extraneous poles are generally easily isolated in the LPC analysis since their bandwidths are often very large, compared to what one would expect for bandwidths typical of speech formants.”

However, they also state:

“...Another difficulty with the analysis is that although the bandwidth of the root is readily determined, it is not generally clear how it is related to actual formant bandwidth. This is because the bandwidth of the root has been shown to be sensitive to the frame duration, frame position, and method of analysis.”

The other method of estimating first formant bandwidth ($B_1$) is through the decay rate $\alpha$ of the exponentially damped sinusoid corresponding to $F_1$ of the pitch period, since $B_1 = \alpha/\pi$ (cf. Hanson, 1997). But this method can not be used with phonation types like breath or whisper because of the random characteristics of these waveforms.

Because linear prediction is very complex computationally and yet does not provide a bandwidth estimate directly related to the actual one, a more neurally plausible method of approximating $Q_1$ is proposed. Examination of the typical energy values of the 330 auditory filters shows that speech spectra are too smoothed for the half power points of $B_1$ to be of use. On the other hand, a peak energy factor $PE_1$, with properties similar to...
Q1, can be rather easily determined. If $E_i$ is the energy of the auditory filter $i$ calculated by squaring and then averaging its output over the 39.37 ms sliding rectangular window (cf. Section 4.4.3 above), then the peak energy value $PE_i$ at $i$ is defined by

$$PE_i = \frac{E_i}{\sqrt{E_{i-8}E_{i+8}}}.$$  \hspace{1cm} (4.10)

As there are 1/48 octave steps in the frequency range of interest (> 101.59 Hz), $i-8$ indicates a frequency 1/6 octave below $i$, and $i+8$ a frequency 1/6 octave above $i$. The denominator is the geometric mean of the energies of $E_{i-8}$ and $E_{i+8}$. Once $PE_i$ has been computed over the entire F1 range, then the maximum value $PE_i$ is found. The frequency corresponding to $i$ of max $PE_i$ is considered the best estimate of F1 while the logarithm of max $PE_i$ is defines the peak energy factor PE1 ($10\log_{10} \text{max } PE_i$) of the first formant. The F1 estimate based on PE1 has the advantage of being far less susceptible to low frequency errors than one based on the peak amplitude A1, as PE1 is fairly insensitive to the strong downward spectral tilt of voiced sounds. Furthermore, the peak energy factor PE1, like the quality factor Q, is dimensionless, invariant with respect to multiplicative frequency shifts, and indicates the amplification at resonance. The total frequency span is set at one-third octave (2 × 1/6 octaves) because

1) a smaller total frequency span yields excessively low values of PE1,

2) one-sixth octave (12.25%) is on the order of a critical band (16% at 1 kHz, Zwicker and Fastl, 1999, Chapter 6; see also Schreiner and Langner, 1997 for a physiological critical band of 0.22 octave or 16.47%).

The upper and lower F1 limits comprise ten one-sixth octaves, with the F1s of women and children set respectively one-fourth and one-half octave higher than men’s:

1) Men 256.00-812.74 Hz (P&B 270-730 Hz; H range 305-963 Hz)
2) Women 304.43-966.52 Hz (P&B 310-860 Hz; H range 331-1163 Hz)

3) Children 362.03-1149.40 Hz (P&B 370-1030 Hz; H range 344-1316 Hz)

The formant means of Peterson and Barney (1952) are indicated by P&B, the formant ranges of Hillenbrand et al. (1995) by H. As with pitch estimation, the upper and lower F1 bounds may be adjusted from their default values to correct a gross estimation error. In the next chapter, signals with known 3 dB bandwidths are synthesized in order to compare PE1 with the Q1 obtained by the linear predictive method.

4.6 Phonetically homogeneous subsegments

The purpose of this section is to present a set of explicit rules for manually dividing the speech signal into phonetically homogeneous subsegments. Manual segmentation is still the method of choice for accurate determination of subsegment boundaries, since automatic methods are evaluated by using the results of hand analysis as the reference standard (Andre-Obrecht, 1988; Kemp et al., 2000). Subsegment boundaries are marked by means of four parameters:

1) Subsegment boundary at the transition from [–speech] to [+speech], and vice versa (Section 4.2.3)

2) Subsegment boundary at the transition from [–voice] to [+voice], and vice versa (Section 4.2.3)

3) Subsegment boundary at the transition from rising to falling F0 as well as from falling to rising F0. When a subsegment is associated with only simple F0 contours, i.e. rising, falling, or level, then the contour can be fully specified by its
A subsegment boundary is also placed at an abrupt shift in pitch register, i.e. between creaky voice, modal voice, or falsetto.

4) Subsegment boundary at a local maximum in the spectral distance function between adjacent frames (metric-based segmentation, Kemp et al., 2000). The between-frame spectral distance is determined in much the same manner as it is in Section 4.4.3. The Euclidean distance is calculated between two normalized 330-element speech energy vectors (NSE) separated by the time step of 2.46 ms. The distance function of the signal is scanned to obtain the maximum value, then all the values are normalized relative to the maximum value and multiplied by 100. Typically, the spectral distance function of a subsegment is U-shaped, with peaks occurring at the beginning and end of the subsegment, and a valley near its midpoint.

When there is a conflict amongst the four criteria, the higher ranking one takes precedence. For example, a subsegment boundary between a sharp rise and fall in F0 according to 3), is preferred over one marked by a nearby peak in the between-frame spectral distance function according to 4).

Each subsegment is sampled at three temporal locations: initial, mid, and final. The initial and final locations are those frames which are, respectively, closest to 1/8 (12.5%) and 7/8 (87.5%) of the duration of the subsegment. The rationale behind this choice is twofold. First, the very beginnings and ends of subsegments are often characterized by a great deal of variability in pitch and formant values. On the other hand, those values are reasonably stable at 1/8 and 7/8 of the duration of the subsegment. Second, one-eighth of the duration of the subsegment is in the range of the Weber
fraction for duration discrimination, 10% to 20% depending on the method used (see Boyle and Espinoza-Varas, 1990, 1992 for fricatives and affricates as well as Schlauch et al., 2001 for additional references). The mid temporal location is either

1) the frame with the maximum energy or

2) the frame nearest the halfway point (50% of subsegment duration) on the condition that the frame with the maximum energy is found to be at a point less than 1/8 or greater than 7/8 of the duration of the subsegment.

The first case is typical of sonorants, and vowels in particular, where the frame with maximum energy usually lies somewhere between the initial and final frames. The second case is typical of obstruents, where the frame with maximum energy often occurs at the very beginning (postvocalic) or at the very end (prevocalic) of the subsegment. The frame with the maximum energy is chosen because it yields the best environmental signal-to-noise ratio.

4.7 Summary

This chapter describes the different speech processing algorithms used in the present study. A bank of second-order linear filters roughly models the response of the auditory periphery. The autocorrelation of the sum of the low-frequency filter outputs yields estimates of F0 and modulation noise (log pACC). The narrow-band slope (H1–H2) is calculated automatically on the basis of this F0 estimate. The aspiration noise (HNR) and the wide-band harmonic slope, are computed simultaneously by means of the ASHE least squares table lookup procedure (Analysis-by-Synthesis Harmonic
Estimation). A method of approximating the Q of the first formant (PE1) is proposed as well as a set of segmentation and sampling criteria.

In the following chapter, the proposed speech processing methods are evaluated mainly by means of synthesized source signals.
NOTES TO CHAPTER 4

1. The theoretical relation $B_{3\text{dB}} = (2 / \pi)ERB$ or $0.636ERB$ in Equation 4.4 is discussed by Hartmann (1998: 262-263). Lopez-Poveda and Meddis (2001) modeled the auditory filter with pulsation threshold data and found the experimental result $B_{3\text{dB}} \approx 0.89ERB$. If the latter relation were adopted, then $Q_{max_{10\text{dB}}} \approx 3.47$, a somewhat small value in light of the psychophysical and physiological evidence.

2. The relation $Q_{10\text{dB}} \approx Q_{3\text{dB}} / 3$ is valid if the high Q (and thus fairly symmetrical) resonance curve is approximated by an equivalent lowpass filter whose positive and negative frequency response is shifted upward in frequency. The squared magnitude response of a one pole lowpass filter is $1/(1+(f / f_c)^2)$, where $f$ is the frequency variable and $f_c$ the 3 dB cutoff frequency. The response of the filter at $f$ is $-3 \text{ dB}$ when $f$ is equal to $f_c$ and $-10 \text{ dB}$ when $f$ is equal to $3f_c$. As $B_{3\text{dB}} = 2f_c$, $B_{3\text{dB}}$ must be multiplied by 3 to obtain $B_{10\text{dB}}$. Hence the $Q_{10\text{dB}}$ of the equivalent filter is $Q_{3\text{dB}} / 3$, given $Q_{3\text{dB}} = f_c / B_{3\text{dB}}$ and $Q_{10\text{dB}} = f_c / B_{10\text{dB}}$.

3. A number of pitch detection algorithms, including the autocorrelation method, preprocess the speech signal using a lowpass filter with a cutoff frequency of 900 Hz. (Rabiner et al., 1976; Hess, 1983).

4. The lowest note on the piano keyboard has a nominal value of 27.5 Hz (A0).

5. Note also that the power spectrum of a signal is the Fourier transform of its autocorrelation function (Rabiner and Schafer, 1978: 176-179). Division of the
autocorrelation function by a constant does not alter the frequency distribution of the signal.

6. The synthesized noise is the sum of equal amplitude sinusoids with equiprobable random phase spaced 0.1 Hz apart.

7. I wish to thank James Hillenbrand for making all his data on American English vowels, including audio files and statistical analysis, available to me.

8. The British tradition of intonation analysis considers the simple rising or falling pitch contours to be unitary (see Ladd, 1980, Chapter I for a review). There have been various attempts to assign abstract paralexical meanings to rising or falling contours. For example, Cruttenden (1986: 168) examined the cross-linguistic evidence and observed that there are nearly universal differences between the meanings of rising and falling intonations.

<table>
<thead>
<tr>
<th>Falling</th>
<th>Rising</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutral statement</td>
<td>Implicational or tentative statement; yes/no question</td>
</tr>
<tr>
<td>Sentence final</td>
<td>Sentence non-final</td>
</tr>
<tr>
<td>Neutral question word question</td>
<td>Sympathetic question word question</td>
</tr>
<tr>
<td>Command</td>
<td>Request</td>
</tr>
</tbody>
</table>

These intonational meanings are associated with the simple contours regardless of their starting and ending pitch levels, which is a strong indication that the simple contours function as a linguistic unit. In addition, more recent phonological work on lexically contrastive tone (e.g. Bao, 1999) has demonstrated that simple rising or falling contours represent elementary tonal units in a number of languages and do not consist merely of a sequence of a low and high tone or a high tone and low tone, respectively. Note that the
existence of unit contours does not exclude the use of different pitch registers (see Yip 2002: 52-56 for further discussion and references).
CHAPTER 5

EVALUATION OF THE SPEECH ANALYSIS METHODS

5.1 Assessment of pitch and voicing decisions

5.1.1 Performance at very low fundamental frequencies

Rabiner and Schafer (1978: 145) state that “to get any indication of periodicity in the autocorrelation function, the window must have a duration of at least two periods of the waveform.” Accordingly, the 39.37 ms rectangular window of the present study should yield F0 estimates only down to 50.8 Hz; a 78.7 ms window would be required for F0 estimates down to 25.4 Hz. The latter fundamental frequency lies somewhat below the lower limit of melodic pitch (see Section 4.2.2). To illustrate the behavior of the autocorrelation function with the 39.37 ms window, three sawtooth waves 2.5 s long are synthesized at a sampling rate of 44.1 kHz. A sawtooth is a reasonably good approximation to the differentiated glottal wave, since its harmonic spectrum falls off at a theoretical rate of –6 dB/octave. The F0s are fixed at 27.5 Hz, 38.9 Hz (half octave), and 55 Hz. The fundamental period of the autocorrelation function is searched over the range 26.14-285.28 Hz using 1) the value of the highest autocorrelation coefficient and 2) the value given by the peak picking method described in Section 4.2.2. The means and standard deviations of the three F0s are calculated and the results of the two methods are compared in Table 5.1. As expected, the maximum value method exhibits extremely poor performance at 27.5 Hz and 38.9 Hz. The peak picking method, on the other hand, shows far superior estimation ability at the same fundamental frequencies. Consequently, the
proposed peak picking method allows meaningful autocorrelation pitch extraction even when the window has a duration of about one period of the waveform. Hence the time resolution of the autocorrelation function is improved by a factor of two.

As the fundamental frequency declines, the log pACC increasingly dips below the absolute voicing threshold of –9 dB during the small amplitude portions of the waveform (cf. the decreasing log pACC means at 55 Hz, 38.9 Hz, and 27.5 Hz: –2.76 dB, –5.28 dB, –11.64 dB). To account for such cases, the program supplies a set of unthresholded pitch estimates. Another point of interest is how well the ASHE procedure (Section 4.4) assesses the mid to high frequency tilt of these low fundamentals. The means and standard deviations (in parentheses) of the harmonic slopes at 55 Hz, 38.9 Hz, and 27.5 Hz are respectively –6.09 dB/octave (0.19), –6.06 dB/octave (0.42), and –6.10 dB/octave (0.22). Thus there is excellent agreement between the predicted and measured values.

<table>
<thead>
<tr>
<th>Synthesized F0</th>
<th>27.5 Hz</th>
<th>38.9 Hz</th>
<th>55 Hz</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Measured values (Hz)</strong></td>
<td>mean</td>
<td>s.d.</td>
<td>mean</td>
</tr>
<tr>
<td>Maximum value method</td>
<td>281.19</td>
<td>0.00</td>
<td>179.54</td>
</tr>
<tr>
<td>Peak peaking method</td>
<td>31.52</td>
<td>2.67</td>
<td>41.30</td>
</tr>
</tbody>
</table>

Table 5.1 Means and standard deviations of the F0s of sawtooth waves with a 39.37 ms rectangular window. The maximum value method finds the highest value of the autocorrelation function. The proposed peak picking method is described in Section 4.2.2.
5.1.2 Pitch and voicing errors

As a test of the pitch and voicing determination methods, error rates are determined by means of four tokens of one sentence from the TIMIT corpus: “She had your dark suit in greasy wash water all year.” The tokens were uttered by two men and two women and recorded at a sampling rate of 16 kHz. Two pitch trackers provide the analysis, the one described in the last chapter and the WaveSurfer program (K. Sjölander and J. Beskow, version 1.6.2). The program includes the well-known ESPS algorithm (get_f0) written by D. Lin and revised by D. Talkin. A description of the ESPS algorithm can be found in Talkin (1995). The pitch and voicing estimates are obtained through a

<table>
<thead>
<tr>
<th>Procedure</th>
<th>VU Errors</th>
<th>UV Errors</th>
<th>Voicing Errors</th>
<th>High GP Errors</th>
<th>Low GP Errors</th>
<th>GP Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Men</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ESPS method</td>
<td>5.95%</td>
<td>1.87%</td>
<td>7.82%</td>
<td>0.00%</td>
<td>1.06%</td>
<td>1.06%</td>
</tr>
<tr>
<td>Present AC method</td>
<td>5.27%</td>
<td>0.60%</td>
<td>5.87%</td>
<td>2.82%</td>
<td>0.47%</td>
<td>3.29%</td>
</tr>
<tr>
<td><strong>Women</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ESPS method</td>
<td>5.42%</td>
<td>0.59%</td>
<td>6.01%</td>
<td>1.25%</td>
<td>2.61%</td>
<td>3.86%</td>
</tr>
<tr>
<td>Present AC method</td>
<td>1.69%</td>
<td>1.69%</td>
<td>3.38%</td>
<td>1.93%</td>
<td>0.68%</td>
<td>2.61%</td>
</tr>
<tr>
<td><strong>Overall Mean</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ESPS method</td>
<td>6.91%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.46%</td>
</tr>
<tr>
<td>Present AC method</td>
<td>4.62%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.95%</td>
</tr>
</tbody>
</table>

Table 5.2 Voicing and gross pitch (GP) errors using the present autocorrelation (AC) and ESPS methods. The voiced-to-unvoiced (VU), unvoiced-to-voiced (UV), and total voicing errors are tabulated as well as the high, low, and total gross pitch errors.
normalized cross-correlation procedure, and then smoothed by a post-processing stage of
dynamic programming. The male and female F0 ranges (26-285 Hz; 26-480 Hz) are set
to the same default values as in the present autocorrelation method. The voicing decisions
and F0s of both programs are evaluated manually by comparing them frame-by-frame
with visual estimates of the pitch periodicity of the speech waveform. Two types of
inaccuracy are counted: 1) voicing decisions involving voiced-to-unvoiced or unvoiced-
to-voiced errors (VU or UV errors), 2) gross pitch errors that overestimate or
underestimate the F0 by 20% (high or low gross pitch errors).

As shown in Table 5.2, the autocorrelation method leads to fewer overall voicing
errors than the ESPS algorithm (4.62% vs. 6.91%). In addition, the overall gross pitch
errors of the present method are only slightly larger than those of the ESPS algorithm
(2.95% vs. 2.46%) even though no post-processing corrections are applied. It was pointed
out in Section 4.2.2 that the value of the autocorrelation coefficient increases with shorter
lags, thereby weighting the peak toward the higher fundamental frequencies. Thus the
high gross pitch errors are seen to predominate. The cross-correlation coefficient of the
ESPS algorithm should in principle be immune to such preferential weighting effects.
Nevertheless, the low gross pitch errors are the main source of the overall gross pitch
errors in the ESPS method. It is instructive to compare these results with those of
Bagshaw et al. (1993) who employed a normalized cross-correlation technique with a
very similar window duration (38.4 ms). Using the same gross pitch error criterion of
20%, they found an overall voicing error of 14.28% and an overall gross pitch error of
1.06%. Although the voicing error is much larger than those of the two methods treated
In this section, the overall gross pitch error is about the lowest cited in the literature (cf. de Cheveigné and Kawahara, 2002).

In sum, the proposed autocorrelation method shows a substantially fewer number of voicing errors than the ESPS cross-correlation algorithm. There is, however, a slightly greater incidence of gross pitch errors. In addition to the error proclivities inherent in the F0 extraction method, nonlinearities of the vocal fold system can lead to pitch breaks and irregular vibration. In view of these difficulties, Granqvist and Hammarberg (2003) proposed hand-selecting F0 candidates from a visual display of all the correlation lags (correlogram). Nevertheless, this method is very time-consuming. In the present study, the upper and lower F0 bounds are adjusted whenever the first-pass pitch estimates show an erroneous departure from the observed periodicity of the speech waveform. If, for example, there is a high octave jump during creaky voice, the upper F0 bound is lowered to eliminate it.

5.2 The log pACC and synthetic modulation noise

As was noted in Chapter 3, the PA (the pitch amplitude or inverse filtered peak autocorrelation coefficient) is shown to be sensitive to all forms of aperiodicity, whether it arises from irregular patterns of vocal fold vibration (modulation noise) or turbulent noise sources (aspiration noise). The peak autocorrelation coefficient without inverse filtering is hypothesized to present a similar sensitivity. To evaluate the log pACC as an acoustic measure of vocal source irregularity, signals are constructed with varying amounts of synthetic modulation noise (jitter and shimmer). Each signal consists of a unit pulse train sampled at 44.1 kHz, a quantization fine enough to avoid a significant
artefactual increase in the measured jitter (Horii, 1979; Titze et al., 1987; Schoentgen and De Guchteneere, 1991), but coarse enough to make the resulting log pACC values comparable to those of speech at standard sampling rates (≤ 44.1 kHz). The periods and amplitudes are modulated on a cycle-to-cycle basis by a pseudo-Gaussian random variable normalized so that the mean is zero and the standard deviation one. To fix the level of percent jitter or shimmer in dB, the normalized variable is multiplied by the corresponding equivalent fraction. The modulated pulse train is delivered to a lowpass digital filter (equivalent to a RC circuit) with a magnitude function falling off at –6 dB/octave

$$H(z) = \frac{1 - e^{-2\pi f_c T}}{1 - e^{-2\pi f_c T} z^{-1}}. \quad (5.1)$$

The sampling rate $T$ is 44.1 kHz while the –3 dB cutoff frequency $f_c$ of the filter is fixed at 0.75×F0. This cutoff frequency insures that the falling exponential branch of the waveform is sufficiently close to zero before the next impulse occurs. The waveform thus constitutes a highly simplified model of the differentiated glottal pulse (see Figure 5.1).
Four F0s (72.88, 128.93, 229.72, 273.19 Hz) are generated with ten levels of jitter (0.5 to 5.0% in 0.5% steps) and ten levels of shimmer (0.19 to 3.57 dB in 0.376 dB steps). The middle F0 values, 128.93 and 229.72 Hz, lie near the central tendency of the adult male and female pitch ranges (Peterson and Barney, 1952; Baken and Orlikoff, 2000). The synthesized jitter ranges from the usual levels of normal voice (≤1%) to 5%, the largest value of harsh creaky voice observed in normal men (Horii, 1985). The synthesized shimmer ranges from levels typical of normal voice (≤0.5 dB) to 3.57 dB, somewhat above the maximum value of 3.23 dB found by Kitajima and Gould (1976) for pathological voice.

Analysis of the 2.5 s signals yields measures of the parameters F0, log pACC, and HNR. To avoid octave jumps at elevated jitter and shimmer levels, the upper and lower F0 bounds were set to be ±1/2 octave of the fundamental frequency. The synthetic signals occasionally produce frames in which the pACC is nearly zero, particularly when the F0 is low and the shimmer level high. As these outliers would introduce important biases into the mean, only those frames with a log pACC greater than –96.33 dB are included in the calculations. The means and standard deviations of the parameters are presented in Tables 5.3.1 through 5.3.4 below. To facilitate comparison, the mean log pACC is plotted against the synthesized jitter and shimmer in Figure 5.2.

As expected from the discussion in Section 4.2.2, the lower the fundamental frequency, the lower the log pACC, except at the highest shimmer levels. The linear regression lines follow quite closely the ten points of the jitter scatter plots. However, the lines show a somewhat poorer fit in the shimmer scatter plots. Accordingly in Table 5.4 below, the Pearson product moment correlation coefficients exhibit strong linear
associations between the log pACC and jitter whereas the associations between the log pACC and shimmer are slightly weaker. These monotonic relationships suggest that the peak autocorrelation coefficient is the most natural and economical parameter for characterizing relative modulation noise since it is already used in the determination of F0 and voicing.

Because jitter is random frequency modulation, it introduces side bands around each of the harmonics (Klingholz and Martin, 1985; Hillenbrand, 1987; Murphy, 2000). However, these same spectral modifications may lead to a decrease in the measured HNR. As discussed in the background chapter, the focus of a number of studies has been to develop techniques that reduce the influence of jitter and shimmer on HNR measures. The ASHE algorithm is likewise seen to be fairly insensitive to jitter and shimmer. Examination of Tables 5.3.1 through 5.3.4 shows that jitter has no measurable effect on the HNR unless the F0 and percent jitter simultaneously exceed 200 Hz and 3.5%.
Figure 5.2 The mean pACCs plotted against synthesized jitter (top) and shimmer (bottom) at four F0s. A regression line is fitted to the ten points of each frequency set.
<table>
<thead>
<tr>
<th>Jitter (%)</th>
<th>Log pACC (dB)</th>
<th>HNR (dB)</th>
<th>F0 (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>s.d.</td>
<td>mean</td>
</tr>
<tr>
<td>0.50</td>
<td>-1.89</td>
<td>0.12</td>
<td>23.33</td>
</tr>
<tr>
<td>1.00</td>
<td>-1.94</td>
<td>0.15</td>
<td>23.33</td>
</tr>
<tr>
<td>1.50</td>
<td>-2.02</td>
<td>0.23</td>
<td>23.33</td>
</tr>
<tr>
<td>2.00</td>
<td>-2.12</td>
<td>0.32</td>
<td>23.33</td>
</tr>
<tr>
<td>2.50</td>
<td>-2.24</td>
<td>0.44</td>
<td>23.32</td>
</tr>
<tr>
<td>3.00</td>
<td>-2.36</td>
<td>0.57</td>
<td>23.31</td>
</tr>
<tr>
<td>3.50</td>
<td>-2.50</td>
<td>0.70</td>
<td>23.30</td>
</tr>
<tr>
<td>4.00</td>
<td>-2.64</td>
<td>0.84</td>
<td>23.30</td>
</tr>
<tr>
<td>4.50</td>
<td>-2.79</td>
<td>0.97</td>
<td>23.28</td>
</tr>
<tr>
<td>5.00</td>
<td>-2.93</td>
<td>1.09</td>
<td>23.27</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Shimmer (dB)</th>
<th>Log pACC (dB)</th>
<th>HNR (dB)</th>
<th>F0 (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>s.d.</td>
<td>mean</td>
</tr>
<tr>
<td>0.19</td>
<td>-1.87</td>
<td>0.13</td>
<td>23.33</td>
</tr>
<tr>
<td>0.56</td>
<td>-1.92</td>
<td>0.18</td>
<td>23.33</td>
</tr>
<tr>
<td>0.94</td>
<td>-2.01</td>
<td>0.31</td>
<td>23.33</td>
</tr>
<tr>
<td>1.32</td>
<td>-2.18</td>
<td>0.55</td>
<td>23.33</td>
</tr>
<tr>
<td>1.69</td>
<td>-2.44</td>
<td>1.08</td>
<td>23.33</td>
</tr>
<tr>
<td>2.07</td>
<td>-2.73</td>
<td>1.32</td>
<td>23.32</td>
</tr>
<tr>
<td>2.45</td>
<td>-3.12</td>
<td>1.75</td>
<td>23.27</td>
</tr>
<tr>
<td>2.82</td>
<td>-3.56</td>
<td>2.42</td>
<td>23.24</td>
</tr>
<tr>
<td>3.20</td>
<td>-3.77</td>
<td>2.39</td>
<td>23.14</td>
</tr>
<tr>
<td>3.57</td>
<td>-4.21</td>
<td>2.89</td>
<td>23.07</td>
</tr>
</tbody>
</table>

Table 5.3.1 Means and standard deviations (s.d.) of the log pACC, HNR, and F0 across ten synthesized jitter and shimmer levels at 72.88 Hz.
<table>
<thead>
<tr>
<th>Synthetic F0</th>
<th>128.93 Hz</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log pACC (dB)</td>
<td>HNR (dB)</td>
</tr>
<tr>
<td>mean</td>
<td>s.d.</td>
</tr>
<tr>
<td>Jitter (%)</td>
<td></td>
</tr>
<tr>
<td>0.50</td>
<td>-0.97</td>
</tr>
<tr>
<td>1.00</td>
<td>-1.01</td>
</tr>
<tr>
<td>1.50</td>
<td>-1.07</td>
</tr>
<tr>
<td>2.00</td>
<td>-1.16</td>
</tr>
<tr>
<td>2.50</td>
<td>-1.27</td>
</tr>
<tr>
<td>3.00</td>
<td>-1.38</td>
</tr>
<tr>
<td>3.50</td>
<td>-1.51</td>
</tr>
<tr>
<td>4.00</td>
<td>-1.65</td>
</tr>
<tr>
<td>4.50</td>
<td>-1.79</td>
</tr>
<tr>
<td>5.00</td>
<td>-1.94</td>
</tr>
<tr>
<td>Shimmer (dB)</td>
<td></td>
</tr>
<tr>
<td>0.19</td>
<td>-0.96</td>
</tr>
<tr>
<td>0.56</td>
<td>-1.00</td>
</tr>
<tr>
<td>0.94</td>
<td>-1.08</td>
</tr>
<tr>
<td>1.32</td>
<td>-1.22</td>
</tr>
<tr>
<td>1.69</td>
<td>-1.41</td>
</tr>
<tr>
<td>2.07</td>
<td>-1.66</td>
</tr>
<tr>
<td>2.45</td>
<td>-2.00</td>
</tr>
<tr>
<td>2.82</td>
<td>-2.22</td>
</tr>
<tr>
<td>3.20</td>
<td>-2.87</td>
</tr>
<tr>
<td>3.57</td>
<td>-3.42</td>
</tr>
</tbody>
</table>

Table 5.3.2 Means and standard deviations (s.d.) of the log pACC, HNR, and F0 across ten synthesized jitter and shimmer levels at 128.93 Hz.
<table>
<thead>
<tr>
<th>Jitter (%)</th>
<th>Log pACC (dB)</th>
<th>229.72 Hz</th>
<th>HNR (dB)</th>
<th>F0 (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>s.d.</td>
<td>mean</td>
<td>s.d.</td>
</tr>
<tr>
<td>0.50</td>
<td>-0.52</td>
<td>0.01</td>
<td>23.33</td>
<td>0.00</td>
</tr>
<tr>
<td>1.00</td>
<td>-0.56</td>
<td>0.03</td>
<td>23.33</td>
<td>0.00</td>
</tr>
<tr>
<td>1.50</td>
<td>-0.61</td>
<td>0.05</td>
<td>23.33</td>
<td>0.00</td>
</tr>
<tr>
<td>2.00</td>
<td>-0.69</td>
<td>0.10</td>
<td>23.30</td>
<td>0.23</td>
</tr>
<tr>
<td>2.50</td>
<td>-0.79</td>
<td>0.14</td>
<td>23.13</td>
<td>0.96</td>
</tr>
<tr>
<td>3.00</td>
<td>-0.91</td>
<td>0.20</td>
<td>22.48</td>
<td>2.02</td>
</tr>
<tr>
<td>3.50</td>
<td>-1.04</td>
<td>0.27</td>
<td>21.17</td>
<td>3.09</td>
</tr>
<tr>
<td>4.00</td>
<td>-1.19</td>
<td>0.35</td>
<td>19.27</td>
<td>3.64</td>
</tr>
<tr>
<td>4.50</td>
<td>-1.36</td>
<td>0.42</td>
<td>17.50</td>
<td>3.76</td>
</tr>
<tr>
<td>5.00</td>
<td>-1.54</td>
<td>0.51</td>
<td>16.07</td>
<td>3.65</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Shimmer (dB)</th>
<th>Log pACC (dB)</th>
<th>229.72 Hz</th>
<th>HNR (dB)</th>
<th>F0 (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>s.d.</td>
<td>mean</td>
<td>s.d.</td>
</tr>
<tr>
<td>0.19</td>
<td>-0.52</td>
<td>0.02</td>
<td>23.33</td>
<td>0.00</td>
</tr>
<tr>
<td>0.56</td>
<td>-0.57</td>
<td>0.06</td>
<td>23.33</td>
<td>0.00</td>
</tr>
<tr>
<td>0.94</td>
<td>-0.68</td>
<td>0.12</td>
<td>23.33</td>
<td>0.00</td>
</tr>
<tr>
<td>1.32</td>
<td>-0.86</td>
<td>0.22</td>
<td>23.33</td>
<td>0.00</td>
</tr>
<tr>
<td>1.69</td>
<td>-1.10</td>
<td>0.36</td>
<td>23.33</td>
<td>0.00</td>
</tr>
<tr>
<td>2.07</td>
<td>-1.42</td>
<td>0.57</td>
<td>23.33</td>
<td>0.00</td>
</tr>
<tr>
<td>2.45</td>
<td>-1.83</td>
<td>0.88</td>
<td>23.31</td>
<td>0.25</td>
</tr>
<tr>
<td>2.82</td>
<td>-2.28</td>
<td>1.19</td>
<td>23.23</td>
<td>0.67</td>
</tr>
<tr>
<td>3.20</td>
<td>-2.83</td>
<td>1.58</td>
<td>23.03</td>
<td>1.22</td>
</tr>
<tr>
<td>3.57</td>
<td>-3.45</td>
<td>2.06</td>
<td>22.67</td>
<td>1.76</td>
</tr>
</tbody>
</table>

Table 5.3.3 Means and standard deviations (s.d.) of the log pACC, HNR, and F0 across ten synthesized jitter and shimmer levels at 229.72 Hz.
<table>
<thead>
<tr>
<th>Jitter (%)</th>
<th>Log pACC (dB)</th>
<th>HNR (dB)</th>
<th>F0 (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>s.d.</td>
<td>mean</td>
</tr>
<tr>
<td>0.50</td>
<td>-0.43</td>
<td>0.01</td>
<td>23.33</td>
</tr>
<tr>
<td>1.00</td>
<td>-0.46</td>
<td>0.02</td>
<td>23.33</td>
</tr>
<tr>
<td>1.50</td>
<td>-0.50</td>
<td>0.04</td>
<td>23.33</td>
</tr>
<tr>
<td>2.00</td>
<td>-0.56</td>
<td>0.07</td>
<td>23.28</td>
</tr>
<tr>
<td>2.50</td>
<td>-0.64</td>
<td>0.12</td>
<td>22.69</td>
</tr>
<tr>
<td>3.00</td>
<td>-0.74</td>
<td>0.17</td>
<td>21.36</td>
</tr>
<tr>
<td>3.50</td>
<td>-0.84</td>
<td>0.22</td>
<td>19.67</td>
</tr>
<tr>
<td>4.00</td>
<td>-0.97</td>
<td>0.29</td>
<td>17.53</td>
</tr>
<tr>
<td>4.50</td>
<td>-1.10</td>
<td>0.35</td>
<td>15.83</td>
</tr>
<tr>
<td>5.00</td>
<td>-1.25</td>
<td>0.42</td>
<td>14.33</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Shimmer (dB)</th>
<th>Log pACC (dB)</th>
<th>HNR (dB)</th>
<th>F0 (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>s.d.</td>
<td>mean</td>
</tr>
<tr>
<td>0.19</td>
<td>-0.43</td>
<td>0.01</td>
<td>23.33</td>
</tr>
<tr>
<td>0.56</td>
<td>-0.48</td>
<td>0.05</td>
<td>23.33</td>
</tr>
<tr>
<td>0.94</td>
<td>-0.58</td>
<td>0.10</td>
<td>23.33</td>
</tr>
<tr>
<td>1.32</td>
<td>-0.75</td>
<td>0.19</td>
<td>23.33</td>
</tr>
<tr>
<td>1.69</td>
<td>-0.97</td>
<td>0.31</td>
<td>23.33</td>
</tr>
<tr>
<td>2.07</td>
<td>-1.26</td>
<td>0.47</td>
<td>23.32</td>
</tr>
<tr>
<td>2.45</td>
<td>-1.61</td>
<td>0.67</td>
<td>23.18</td>
</tr>
<tr>
<td>2.82</td>
<td>-2.02</td>
<td>0.95</td>
<td>22.86</td>
</tr>
<tr>
<td>3.20</td>
<td>-2.52</td>
<td>1.36</td>
<td>22.03</td>
</tr>
<tr>
<td>3.57</td>
<td>-3.10</td>
<td>1.98</td>
<td>20.80</td>
</tr>
</tbody>
</table>

Table 5.3.4 Means and standard deviations (s.d.) of the log pACC, HNR, and F0 across ten synthesized jitter and shimmer levels at 273.19 Hz.
Jitter levels greater than 3.5% occur rather infrequently in non-pathological speech. For creaky voice in normal men, Horii (1985) found a range of 0.87-5.05%, with a mean of 2.47% and a standard deviation of 0.92%. The average fundamental frequency, however, was 64.8 Hz (s.d. 12.0), well below the 200 Hz limit. Shimmer appears to have no appreciable effect on the HNR within the modeled range of 0 to 24 dB. Compared to the pitch-scaled harmonic filter method (Jackson and Shadle, 2001), the HNR estimates of the present ASHE technique demonstrate equal insensitivity to shimmer and less sensitivity to jitter at low fundamental frequencies. On the other hand, the HNR of pitch-synchronous harmonic analysis (Murphy, 1999) is degraded by jitter and shimmer to a far smaller degree than either the present technique or the Jackson and Shadle method. As in this study, signals synthesized with the same level of jitter show a decline in the measured HNR with increasing fundamental frequencies. The main drawback to pitch-synchronous analysis is its neural implausibility in view of the precise period-by-period segmentation required.

<table>
<thead>
<tr>
<th>Synthetic F0</th>
<th>72.88 Hz</th>
<th>128.92 Hz</th>
<th>229.72 Hz</th>
<th>273.18 Hz</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson correlation coefficient of the mean log pACC (dB) and synthetic jitter (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.993</td>
<td>&lt;0.001</td>
<td>-0.989</td>
<td>&lt;0.001</td>
<td>-0.980</td>
</tr>
<tr>
<td>Pearson correlation coefficient of the mean log pACC (dB) and synthetic shimmer (dB)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.977</td>
<td>&lt;0.001</td>
<td>-0.950</td>
<td>&lt;0.001</td>
<td>-0.962</td>
</tr>
</tbody>
</table>

**Table 5.4** Pearson product moment correlation coefficient of the differentiated glottal wave between the mean logarithmic peak autocorrelation coefficient (10log₁₀PACC) and synthesized random jitter (%) and shimmer (dB) at four F0s. All correlation coefficients are significant at less than the 0.001 level. (N=10).
5.3 Simultaneous jitter and shimmer

To evaluate the effects of simultaneously occurring jitter and shimmer on the log pACC, the jitter and shimmer of the differentiated glottal wave are set at the average levels of normal voice and harsh creaky voice. Following Horii (1985), the mean jitter and shimmer values of normal voice are 0.87%/0.48 dB whereas those of harsh creaky voice are 2.47%/1.15 dB. These are approximated as 1.0%/0.56 dB and 2.5%/1.32 dB.

The glottal waveforms are synthesized at four fundamental frequencies (72.88, 128.93, 229.72, 273.19 Hz) and the results are given in Table 5.5. When a comparison is made between these values and those of Tables 5.3.1-5.3.4, it is immediately apparent that simultaneous jitter and shimmer produces only an incremental decrease in the log pACC relative to the same jitter or shimmer levels generated alone. For example, at 128.93 Hz, simultaneous jitter and shimmer levels of 2.5%/1.32 dB yield a log pACC of −1.46 dB. Each considered separately, a jitter level of 2.5% leads to a log pACC of −1.27 dB while

<table>
<thead>
<tr>
<th>Jitter (%)</th>
<th>Shimmer (dB)</th>
<th>F0 (Hz)</th>
<th>Log pACC (dB)</th>
<th>HNR (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean s.d.</td>
<td>mean s.d.</td>
<td>mean s.d.</td>
<td>mean s.d.</td>
</tr>
<tr>
<td>1.00</td>
<td>0.56 0.75</td>
<td>73.02</td>
<td>−1.98 0.21</td>
<td>23.33 0.00</td>
</tr>
<tr>
<td>2.50</td>
<td>1.32 1.99</td>
<td>73.27</td>
<td>−2.45 0.72</td>
<td>23.29 0.30</td>
</tr>
<tr>
<td>1.00</td>
<td>0.56 0.46</td>
<td>129.03</td>
<td>−1.04 0.11</td>
<td>23.33 0.00</td>
</tr>
<tr>
<td>2.50</td>
<td>1.32 1.29</td>
<td>129.29</td>
<td>−1.46 0.42</td>
<td>23.33 0.00</td>
</tr>
<tr>
<td>1.00</td>
<td>0.56 0.00</td>
<td>229.72</td>
<td>−0.61 0.08</td>
<td>23.33 0.00</td>
</tr>
<tr>
<td>2.50</td>
<td>1.32 1.16</td>
<td>229.93</td>
<td>−1.11 0.38</td>
<td>22.82 1.47</td>
</tr>
<tr>
<td>1.00</td>
<td>0.56 1.27</td>
<td>273.65</td>
<td>−0.52 0.07</td>
<td>23.33 0.00</td>
</tr>
<tr>
<td>2.50</td>
<td>1.32 1.83</td>
<td>274.33</td>
<td>−1.00 0.32</td>
<td>21.95 2.40</td>
</tr>
</tbody>
</table>

**Table 5.5** Means and standard deviations (s.d.) of the F0, log pACC, and HNR of the differentiated glottal wave, measured at the average jitter and shimmer levels of normal voice (1.0%, 0.56 dB) and harsh voice (2.5%, 1.32 dB) at four synthetic F0s: 72.88, 128.93, 229.72, 273.19 Hz.
the shimmer level of 1.32 dB results in a log pACC of –1.22 dB. Similar outcomes are obtained at the three other fundamental frequencies.

To explore the additional consequences of simultaneous jitter and shimmer on the log pACC, varying jitter levels are synthesized with a constant shimmer pedestal (1.32 dB). Similarly, varying shimmer levels are generated with a constant jitter pedestal (2.5%). In Tables 5.6.1-5.6.2, the means and standard deviations of the log pACC are provided at two fundamental frequencies 128.93 Hz and 229.72 Hz. The data are illustrated in Figure 5.3. Like the log pACCs of the differentiated glottal wave without jitter and shimmer pedestals (cf. Figure 5.2), the regression lines display linear relations for varying jitter and somewhat less linearity for varying shimmer. The very strong Pearson correlation coefficients in Table 5.7 confirm these observations. Note also that a combination of jitter and shimmer affects the HNR predominately at the higher fundamental frequencies, as does jitter alone. The measured HNRs in Table 5.5 are lowered only negligibly when the F0 lies above 200 Hz and the simultaneous jitter and shimmer level is set at 2.5%/1.32 dB.
Figure 5.3 The mean pACCs of the differentiated glottal wave plotted against synthesized jitter on a shimmer pedestal of 1.32 dB (top) and shimmer on a jitter pedestal of 2.5% (bottom) at two F0s. A regression line is fitted to the ten points of each frequency set.
<table>
<thead>
<tr>
<th>Jitter (%)</th>
<th>Log pACC (dB) mean</th>
<th>s.d.</th>
<th>HNR (dB) mean</th>
<th>s.d.</th>
<th>F0 (Hz) mean</th>
<th>s.d.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.50</td>
<td>-1.21</td>
<td>0.28</td>
<td>23.33</td>
<td>0.00</td>
<td>129.97</td>
<td>0.26</td>
</tr>
<tr>
<td>1.00</td>
<td>-1.22</td>
<td>0.30</td>
<td>23.33</td>
<td>0.00</td>
<td>129.03</td>
<td>0.46</td>
</tr>
<tr>
<td>1.50</td>
<td>-1.30</td>
<td>0.33</td>
<td>23.33</td>
<td>0.00</td>
<td>129.15</td>
<td>0.81</td>
</tr>
<tr>
<td>2.00</td>
<td>-1.37</td>
<td>0.38</td>
<td>23.33</td>
<td>0.00</td>
<td>129.27</td>
<td>1.06</td>
</tr>
<tr>
<td>2.50</td>
<td>-1.46</td>
<td>0.42</td>
<td>23.33</td>
<td>0.00</td>
<td>129.29</td>
<td>1.29</td>
</tr>
<tr>
<td>3.00</td>
<td>-1.57</td>
<td>0.48</td>
<td>23.30</td>
<td>0.25</td>
<td>129.27</td>
<td>1.61</td>
</tr>
<tr>
<td>3.50</td>
<td>-1.68</td>
<td>0.54</td>
<td>23.21</td>
<td>0.60</td>
<td>129.36</td>
<td>2.03</td>
</tr>
<tr>
<td>4.00</td>
<td>-1.80</td>
<td>0.60</td>
<td>23.08</td>
<td>0.89</td>
<td>129.40</td>
<td>2.36</td>
</tr>
<tr>
<td>4.50</td>
<td>-1.92</td>
<td>0.68</td>
<td>22.79</td>
<td>1.31</td>
<td>129.50</td>
<td>2.71</td>
</tr>
<tr>
<td>5.00</td>
<td>-2.06</td>
<td>0.75</td>
<td>22.43</td>
<td>1.78</td>
<td>129.57</td>
<td>3.16</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Shimmer (dB)</th>
<th>Jitter pedestal: 2.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.19</td>
<td>-1.23</td>
</tr>
<tr>
<td>0.56</td>
<td>-1.26</td>
</tr>
<tr>
<td>0.94</td>
<td>-1.34</td>
</tr>
<tr>
<td>1.32</td>
<td>-1.46</td>
</tr>
<tr>
<td>1.69</td>
<td>-1.63</td>
</tr>
<tr>
<td>2.07</td>
<td>-1.87</td>
</tr>
<tr>
<td>2.45</td>
<td>-2.17</td>
</tr>
<tr>
<td>2.82</td>
<td>-2.53</td>
</tr>
<tr>
<td>3.20</td>
<td>-2.93</td>
</tr>
<tr>
<td>3.57</td>
<td>-3.45</td>
</tr>
</tbody>
</table>

Table 5.6.1 Means and standard deviations (s.d.) of the log pACC, HNR, and F0 across ten synthesized jitter and shimmer levels at 128.93 Hz. The shimmer and jitter pedestals of the differentiated glottal wave are respectively 1.32 dB and 2.5%. 
<table>
<thead>
<tr>
<th>Jitter (%)</th>
<th>Shimmer pedestal: 1.32 dB</th>
<th>Jitter pedestal: 2.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.50</td>
<td>-0.86 ± 0.24 dB</td>
<td>-0.78 ± 0.16 dB</td>
</tr>
<tr>
<td>1.00</td>
<td>-0.89 ± 0.26 dB</td>
<td>-0.83 ± 0.19 dB</td>
</tr>
<tr>
<td>1.50</td>
<td>-0.94 ± 0.28 dB</td>
<td>-0.94 ± 0.27 dB</td>
</tr>
<tr>
<td>2.00</td>
<td>-1.01 ± 0.32 dB</td>
<td>-1.11 ± 0.38 dB</td>
</tr>
<tr>
<td>2.50</td>
<td>-1.11 ± 0.38 dB</td>
<td>-1.34 ± 0.50 dB</td>
</tr>
<tr>
<td>3.00</td>
<td>-1.22 ± 0.43 dB</td>
<td>-1.64 ± 0.67 dB</td>
</tr>
<tr>
<td>3.50</td>
<td>-1.34 ± 0.50 dB</td>
<td>-1.80 ± 0.77 dB</td>
</tr>
<tr>
<td>4.00</td>
<td>-1.48 ± 0.58 dB</td>
<td>-1.64 ± 0.58 dB</td>
</tr>
<tr>
<td>4.50</td>
<td>-1.64 ± 0.67 dB</td>
<td>-1.80 ± 0.77 dB</td>
</tr>
<tr>
<td>5.00</td>
<td>-1.80 ± 0.77 dB</td>
<td>-1.80 ± 0.77 dB</td>
</tr>
</tbody>
</table>

Table 5.6.2 Means and standard deviations (s.d.) of the log pACC, HNR, and F0 across ten synthesized jitter and shimmer levels at 229.72 Hz. The shimmer and jitter pedestals of the differentiated glottal wave are respectively 1.32 dB and 2.5%.
Summarizing, the log pACC parameter varies in a nearly linear manner with frequency and amplitude modulation noise levels (jitter and shimmer). The effects of jitter and shimmer on the log pACC are not additive; their combination brings about a relatively small reduction in the log pACC level compared to jitter and shimmer alone. The log pACC parameter also displays a quasi-linear relationship with jitter when the shimmer is held constant at levels typical of harsh creaky voice, and likewise in the converse case. Jitter or simultaneous jitter and shimmer reduce measured HNRs only at the highest modulation noise levels above 200 Hz. The ASHE technique therefore nearly always keeps aspiration noise distinct from modulation noise.

5.4 The HNR and synthetic aspiration noise

5.4.1 Evaluation of the ASHE algorithm using the model signals

To test how well the ASHE procedure performs in determining the harmonics-to-noise ratio, sixteen HNRs are synthesized at four fundamental frequencies (72.88, 128.92, 229.72, 273.18 Hz) with the three harmonic slopes of –3.06, –6.11, and –12.23.
The two harmonic slopes of –3.06 and –6.11 dB/octave delimit the range where most of the speech variation was concentrated in the TIMIT signals; only very rarely was the slope flatter than –3 dB/octave. The model harmonic slope of –12.23 dB/octave is designed to simulate those occurring during breathy voice. All the parameter values are a subset of those used in the synthesis phase of the algorithm. The means and standard deviations of the HNRs generated with the above values are shown in Tables 5.8.1-5.8.6 below. Like the relation between synthetic modulation noise and the log pACC, a nearly one-to-one association exists between the actual and mean HNRs as measured by the algorithm. There is a tendency for the lowest and the highest measured HNRs not to reach their true values, particularly at the two lower F0s with a slope of –3.06 dB/octave. Note also the standard deviations are rather large under these conditions. Nevertheless, the extremely strong Pearson correlation coefficients in Table 5.9 indicate that the overall HNR estimation accuracy of the ASHE algorithm is quite good.

The mean log pACCs produced by the synthetic HNRs are likewise included in Tables 5.8.1-5.8.6. The log pACCs increase with F0, as predicted, and exhibit a leveling off once the synthetic HNR exceeds 9 to 12 dB. At the lower HNRs, the mean log pACCs also tend to decline with increasingly flatter slopes (for example, at 128.93 Hz and HNR of 3.76 dB, –1.10, –1.14, –1.32 dB with the slopes of –12.23, –6.11, –3.06 dB/octave, respectively; at 229.72 Hz and HNR of 3.76 dB, –0.75, –0.81, –1.06 dB). It may be useful to attempt a comparison of these findings with those of Hillenbrand (1987), who pointed out that lower synthetic HNRs lead to higher measured jitter and shimmer levels, even in the complete absence of modulation noise. In his study a synthetic 8 dB HNR at
130 Hz gives rise to 2.61% jitter and 1.17 dB shimmer. In Table 5.8.3, a synthetic HNR of 8.28 dB (128.93 Hz, –6.11 dB/octave) yields a pACC of −1.02 dB. Upon examination of Table 5.3.2, one finds that a log pACC of about −1 dB corresponds to 1.0% jitter and 0.56 dB shimmer, both within the range of normal voice. Thus the pACC appears to be less sensitive to additive noise than the waveform measures of jitter and shimmer. Only when the HNR falls below 4 dB does the additive noise significantly affect the pACC.

This lack of sensitivity to additive noise appears to be the consequence of restricting the input of the autocorrelation function to low frequency regions where the local HNR remains relatively elevated. For instance, the input to the autocorrelation function is the sum of the auditory filter outputs between 13.45 Hz and 830.54 Hz. If a source signal has a low HNR and a steeply sloping harmonic envelope, then the harmonic components above 830.54 Hz should be more immersed in the flat spectrum noise than the components below 830.54 Hz. To illustrate this, the mean log pACC of a 8.28 dB HNR signal (128.93 Hz, –6.11 dB/octave) is computed over several input frequency ranges. When the lower frequency limit is 13.45 Hz and the upper frequency limit increases stepwise through 830.54 Hz, 3227.61 Hz and 6455.22 Hz, the mean log pACC decreases correspondingly: −1.02, −1.08, −1.13. More importantly, when the lower limit is 830.54 Hz and the upper limit is 6455.22 Hz, the mean log pACC is −1.44. This latter result clearly indicates that confining the autocorrelation input range to low frequencies reduces the influence of additive noise on the pACC. Therefore, except at the lowest HNRs, the log pACC functions primarily as a cue for modulation noise, a significant outcome in view of the goal of keeping modulation noise distinct from aspiration noise.
<table>
<thead>
<tr>
<th>Actual HNR (dB)</th>
<th>Measured HNR (dB)</th>
<th>Log pACC (dB)</th>
<th>F0 (Hz)</th>
<th>Slope (dB/octave)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>s.d.</td>
<td>mean</td>
<td>s.d.</td>
</tr>
<tr>
<td>Actual F0 and Slope: 72.88 Hz and –3.06 dB/octave</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.75</td>
<td>2.50</td>
<td>4.02</td>
<td>–2.46</td>
<td>0.39</td>
</tr>
<tr>
<td>2.26</td>
<td>3.32</td>
<td>3.72</td>
<td>–2.29</td>
<td>0.34</td>
</tr>
<tr>
<td>3.76</td>
<td>4.63</td>
<td>3.34</td>
<td>–2.18</td>
<td>0.32</td>
</tr>
<tr>
<td>5.27</td>
<td>6.13</td>
<td>3.36</td>
<td>–2.10</td>
<td>0.30</td>
</tr>
<tr>
<td>6.77</td>
<td>7.71</td>
<td>3.45</td>
<td>–2.04</td>
<td>0.28</td>
</tr>
<tr>
<td>8.28</td>
<td>9.16</td>
<td>3.32</td>
<td>–1.99</td>
<td>0.27</td>
</tr>
<tr>
<td>9.78</td>
<td>10.87</td>
<td>3.68</td>
<td>–1.96</td>
<td>0.27</td>
</tr>
<tr>
<td>11.29</td>
<td>12.37</td>
<td>3.71</td>
<td>–1.94</td>
<td>0.26</td>
</tr>
<tr>
<td>12.79</td>
<td>14.13</td>
<td>4.13</td>
<td>–1.92</td>
<td>0.26</td>
</tr>
<tr>
<td>14.30</td>
<td>15.48</td>
<td>4.09</td>
<td>–1.91</td>
<td>0.26</td>
</tr>
<tr>
<td>15.80</td>
<td>16.47</td>
<td>3.82</td>
<td>–1.91</td>
<td>0.26</td>
</tr>
<tr>
<td>17.31</td>
<td>17.38</td>
<td>3.47</td>
<td>–1.90</td>
<td>0.26</td>
</tr>
<tr>
<td>18.81</td>
<td>18.02</td>
<td>3.33</td>
<td>–1.90</td>
<td>0.26</td>
</tr>
<tr>
<td>20.32</td>
<td>18.87</td>
<td>3.30</td>
<td>–1.89</td>
<td>0.25</td>
</tr>
<tr>
<td>21.82</td>
<td>19.51</td>
<td>3.22</td>
<td>–1.89</td>
<td>0.25</td>
</tr>
<tr>
<td>23.33</td>
<td>19.97</td>
<td>3.11</td>
<td>–1.89</td>
<td>0.25</td>
</tr>
<tr>
<td>Actual F0 and Slope: 128.93 Hz and –3.06 dB/octave</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.75</td>
<td>1.95</td>
<td>2.20</td>
<td>–1.67</td>
<td>0.23</td>
</tr>
<tr>
<td>2.26</td>
<td>2.94</td>
<td>2.54</td>
<td>–1.47</td>
<td>0.18</td>
</tr>
<tr>
<td>3.76</td>
<td>4.31</td>
<td>2.60</td>
<td>–1.32</td>
<td>0.14</td>
</tr>
<tr>
<td>5.27</td>
<td>5.85</td>
<td>2.74</td>
<td>–1.22</td>
<td>0.11</td>
</tr>
<tr>
<td>6.77</td>
<td>7.40</td>
<td>2.88</td>
<td>–1.14</td>
<td>0.09</td>
</tr>
<tr>
<td>8.28</td>
<td>9.08</td>
<td>3.18</td>
<td>–1.09</td>
<td>0.07</td>
</tr>
<tr>
<td>9.78</td>
<td>10.70</td>
<td>3.51</td>
<td>–1.05</td>
<td>0.06</td>
</tr>
<tr>
<td>11.29</td>
<td>12.27</td>
<td>3.79</td>
<td>–1.02</td>
<td>0.05</td>
</tr>
<tr>
<td>12.79</td>
<td>13.68</td>
<td>3.53</td>
<td>–1.00</td>
<td>0.05</td>
</tr>
<tr>
<td>14.30</td>
<td>15.25</td>
<td>3.72</td>
<td>–0.99</td>
<td>0.04</td>
</tr>
<tr>
<td>15.80</td>
<td>16.45</td>
<td>3.59</td>
<td>–0.98</td>
<td>0.04</td>
</tr>
<tr>
<td>17.31</td>
<td>17.82</td>
<td>3.57</td>
<td>–0.97</td>
<td>0.04</td>
</tr>
<tr>
<td>18.81</td>
<td>18.95</td>
<td>3.50</td>
<td>–0.97</td>
<td>0.03</td>
</tr>
<tr>
<td>20.32</td>
<td>19.85</td>
<td>3.32</td>
<td>–0.96</td>
<td>0.03</td>
</tr>
<tr>
<td>21.82</td>
<td>20.66</td>
<td>3.17</td>
<td>–0.96</td>
<td>0.03</td>
</tr>
<tr>
<td>23.33</td>
<td>21.05</td>
<td>3.09</td>
<td>–0.96</td>
<td>0.03</td>
</tr>
</tbody>
</table>

*Table 5.8.1* Means and standard deviations (s.d.) of the measured HNR and log pACC as a function of the sixteen model HNRs generated at 72.88 Hz and 128.93 Hz. The actual harmonic slope is –3.06 dB/slope.
<table>
<thead>
<tr>
<th>Actual HNR (dB)</th>
<th>Measured HNR (dB)</th>
<th>Log pACC (dB)</th>
<th>F0 (Hz)</th>
<th>Slope (dB/octave)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>s.d.</td>
<td>mean</td>
<td>s.d.</td>
</tr>
<tr>
<td>Actual F0 and Slope: 229.72 Hz and –3.06 dB/octave</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.75</td>
<td>1.76</td>
<td>1.90</td>
<td>–1.57</td>
<td>0.28</td>
</tr>
<tr>
<td>2.26</td>
<td>2.76</td>
<td>2.37</td>
<td>–1.28</td>
<td>0.21</td>
</tr>
<tr>
<td>3.76</td>
<td>4.27</td>
<td>2.72</td>
<td>–1.06</td>
<td>0.15</td>
</tr>
<tr>
<td>5.27</td>
<td>5.90</td>
<td>3.01</td>
<td>–0.91</td>
<td>0.12</td>
</tr>
<tr>
<td>6.77</td>
<td>7.61</td>
<td>3.59</td>
<td>–0.79</td>
<td>0.09</td>
</tr>
<tr>
<td>8.28</td>
<td>9.16</td>
<td>3.56</td>
<td>–0.71</td>
<td>0.07</td>
</tr>
<tr>
<td>9.78</td>
<td>10.85</td>
<td>3.94</td>
<td>–0.65</td>
<td>0.05</td>
</tr>
<tr>
<td>11.29</td>
<td>12.35</td>
<td>3.80</td>
<td>–0.61</td>
<td>0.04</td>
</tr>
<tr>
<td>12.79</td>
<td>14.01</td>
<td>4.21</td>
<td>–0.58</td>
<td>0.03</td>
</tr>
<tr>
<td>14.30</td>
<td>14.96</td>
<td>3.76</td>
<td>–0.56</td>
<td>0.03</td>
</tr>
<tr>
<td>15.80</td>
<td>15.91</td>
<td>3.38</td>
<td>–0.55</td>
<td>0.02</td>
</tr>
<tr>
<td>17.31</td>
<td>17.04</td>
<td>3.26</td>
<td>–0.53</td>
<td>0.02</td>
</tr>
<tr>
<td>18.81</td>
<td>18.53</td>
<td>3.24</td>
<td>–0.53</td>
<td>0.02</td>
</tr>
<tr>
<td>20.32</td>
<td>19.74</td>
<td>3.19</td>
<td>–0.52</td>
<td>0.01</td>
</tr>
<tr>
<td>21.82</td>
<td>20.87</td>
<td>2.90</td>
<td>–0.52</td>
<td>0.01</td>
</tr>
<tr>
<td>23.33</td>
<td>21.59</td>
<td>2.68</td>
<td>–0.51</td>
<td>0.01</td>
</tr>
<tr>
<td>Actual F0 and Slope: 273.19 Hz and –3.06 dB/octave</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.75</td>
<td>1.51</td>
<td>1.17</td>
<td>–1.49</td>
<td>0.26</td>
</tr>
<tr>
<td>2.26</td>
<td>2.52</td>
<td>1.70</td>
<td>–1.20</td>
<td>0.19</td>
</tr>
<tr>
<td>3.76</td>
<td>4.03</td>
<td>1.81</td>
<td>–0.99</td>
<td>0.14</td>
</tr>
<tr>
<td>5.27</td>
<td>5.58</td>
<td>1.99</td>
<td>–0.83</td>
<td>0.11</td>
</tr>
<tr>
<td>6.77</td>
<td>7.25</td>
<td>2.32</td>
<td>–0.72</td>
<td>0.08</td>
</tr>
<tr>
<td>8.28</td>
<td>8.90</td>
<td>2.74</td>
<td>–0.63</td>
<td>0.06</td>
</tr>
<tr>
<td>9.78</td>
<td>10.54</td>
<td>3.15</td>
<td>–0.57</td>
<td>0.05</td>
</tr>
<tr>
<td>11.29</td>
<td>12.25</td>
<td>3.45</td>
<td>–0.53</td>
<td>0.04</td>
</tr>
<tr>
<td>12.79</td>
<td>13.88</td>
<td>3.81</td>
<td>–0.50</td>
<td>0.03</td>
</tr>
<tr>
<td>14.30</td>
<td>15.14</td>
<td>3.68</td>
<td>–0.48</td>
<td>0.03</td>
</tr>
<tr>
<td>15.80</td>
<td>16.28</td>
<td>3.45</td>
<td>–0.46</td>
<td>0.02</td>
</tr>
<tr>
<td>17.31</td>
<td>17.04</td>
<td>3.26</td>
<td>–0.53</td>
<td>0.02</td>
</tr>
<tr>
<td>18.81</td>
<td>18.16</td>
<td>3.02</td>
<td>–0.44</td>
<td>0.02</td>
</tr>
<tr>
<td>20.32</td>
<td>19.08</td>
<td>2.86</td>
<td>–0.44</td>
<td>0.02</td>
</tr>
<tr>
<td>21.82</td>
<td>19.85</td>
<td>2.61</td>
<td>–0.43</td>
<td>0.04</td>
</tr>
<tr>
<td>23.33</td>
<td>20.49</td>
<td>2.44</td>
<td>–0.43</td>
<td>0.01</td>
</tr>
</tbody>
</table>

**Table 5.8.2** Means and standard deviations (s.d.) of the measured HNR and log pACC as a function of the sixteen model HNRs generated at 229.72 Hz and 273.19 Hz. The actual harmonic slope is –3.06 dB/slope.
<table>
<thead>
<tr>
<th>Actual HNR (dB)</th>
<th>Measured HNR (dB)</th>
<th>Log pACC (dB)</th>
<th>F0 (Hz)</th>
<th>Slope (dB/octave)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>s.d.</td>
<td>mean</td>
<td>s.d.</td>
</tr>
<tr>
<td>0.75</td>
<td>0.89</td>
<td>0.44</td>
<td>-2.16</td>
<td>0.23</td>
</tr>
<tr>
<td>2.26</td>
<td>2.26</td>
<td>0.58</td>
<td>-2.07</td>
<td>0.21</td>
</tr>
<tr>
<td>3.76</td>
<td>3.79</td>
<td>0.53</td>
<td>-2.01</td>
<td>0.19</td>
</tr>
<tr>
<td>5.27</td>
<td>5.31</td>
<td>0.49</td>
<td>-1.97</td>
<td>0.18</td>
</tr>
<tr>
<td>6.77</td>
<td>6.83</td>
<td>0.46</td>
<td>-1.94</td>
<td>0.17</td>
</tr>
<tr>
<td>8.28</td>
<td>8.34</td>
<td>0.46</td>
<td>-1.92</td>
<td>0.17</td>
</tr>
<tr>
<td>9.78</td>
<td>9.86</td>
<td>0.52</td>
<td>-1.90</td>
<td>0.16</td>
</tr>
<tr>
<td>11.29</td>
<td>11.40</td>
<td>0.60</td>
<td>-1.89</td>
<td>0.16</td>
</tr>
<tr>
<td>12.79</td>
<td>12.90</td>
<td>0.72</td>
<td>-1.89</td>
<td>0.16</td>
</tr>
<tr>
<td>14.30</td>
<td>14.47</td>
<td>0.93</td>
<td>-1.88</td>
<td>0.16</td>
</tr>
<tr>
<td>15.80</td>
<td>16.03</td>
<td>1.24</td>
<td>-1.88</td>
<td>0.16</td>
</tr>
<tr>
<td>17.31</td>
<td>17.63</td>
<td>1.70</td>
<td>-1.87</td>
<td>0.16</td>
</tr>
<tr>
<td>18.81</td>
<td>19.15</td>
<td>1.83</td>
<td>-1.87</td>
<td>0.16</td>
</tr>
<tr>
<td>20.32</td>
<td>20.43</td>
<td>1.82</td>
<td>-1.87</td>
<td>0.16</td>
</tr>
<tr>
<td>21.82</td>
<td>21.42</td>
<td>1.69</td>
<td>-1.87</td>
<td>0.16</td>
</tr>
<tr>
<td>23.33</td>
<td>22.18</td>
<td>1.48</td>
<td>-1.87</td>
<td>0.16</td>
</tr>
<tr>
<td>0.75</td>
<td>0.95</td>
<td>0.51</td>
<td>-1.31</td>
<td>0.14</td>
</tr>
<tr>
<td>2.26</td>
<td>2.34</td>
<td>0.59</td>
<td>-1.21</td>
<td>0.11</td>
</tr>
<tr>
<td>3.76</td>
<td>3.86</td>
<td>0.55</td>
<td>-1.14</td>
<td>0.09</td>
</tr>
<tr>
<td>5.27</td>
<td>5.36</td>
<td>0.53</td>
<td>-1.08</td>
<td>0.08</td>
</tr>
<tr>
<td>6.77</td>
<td>6.87</td>
<td>0.53</td>
<td>-1.04</td>
<td>0.06</td>
</tr>
<tr>
<td>8.28</td>
<td>8.39</td>
<td>0.55</td>
<td>-1.02</td>
<td>0.05</td>
</tr>
<tr>
<td>9.78</td>
<td>9.89</td>
<td>0.59</td>
<td>-1.00</td>
<td>0.05</td>
</tr>
<tr>
<td>11.29</td>
<td>11.42</td>
<td>0.66</td>
<td>-0.98</td>
<td>0.04</td>
</tr>
<tr>
<td>12.79</td>
<td>12.95</td>
<td>0.74</td>
<td>-0.98</td>
<td>0.04</td>
</tr>
<tr>
<td>14.30</td>
<td>14.47</td>
<td>0.88</td>
<td>-0.97</td>
<td>0.03</td>
</tr>
<tr>
<td>15.80</td>
<td>16.11</td>
<td>1.05</td>
<td>-0.96</td>
<td>0.03</td>
</tr>
<tr>
<td>17.31</td>
<td>17.76</td>
<td>1.27</td>
<td>-0.96</td>
<td>0.03</td>
</tr>
<tr>
<td>18.81</td>
<td>19.53</td>
<td>1.53</td>
<td>-0.96</td>
<td>0.03</td>
</tr>
<tr>
<td>20.32</td>
<td>21.26</td>
<td>1.72</td>
<td>-0.96</td>
<td>0.03</td>
</tr>
<tr>
<td>21.82</td>
<td>22.35</td>
<td>1.55</td>
<td>-0.95</td>
<td>0.03</td>
</tr>
<tr>
<td>23.33</td>
<td>22.81</td>
<td>1.26</td>
<td>-0.95</td>
<td>0.02</td>
</tr>
</tbody>
</table>

**Table 5.8.3** Means and standard deviations (s.d.) of the measured HNR and log pACC as a function of the sixteen model HNRs generated at 72.88 Hz and 128.93 Hz. The actual harmonic slope is –6.11 dB/slope.
<table>
<thead>
<tr>
<th>Actual HNR (dB)</th>
<th>Measured HNR (dB)</th>
<th>Log pACC (dB)</th>
<th>F0 (Hz)</th>
<th>Slope (dB/octave)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>s.d.</td>
<td>mean</td>
<td>s.d.</td>
</tr>
<tr>
<td>Actual F0 and Slope: 229.72 Hz and –6.11 dB/octave</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.75</td>
<td>0.96</td>
<td>0.52</td>
<td>–1.08</td>
<td>0.16</td>
</tr>
<tr>
<td>2.26</td>
<td>2.33</td>
<td>0.68</td>
<td>–0.92</td>
<td>0.12</td>
</tr>
<tr>
<td>3.76</td>
<td>3.86</td>
<td>0.62</td>
<td>–0.81</td>
<td>0.09</td>
</tr>
<tr>
<td>5.27</td>
<td>5.37</td>
<td>0.60</td>
<td>–0.72</td>
<td>0.07</td>
</tr>
<tr>
<td>6.77</td>
<td>6.90</td>
<td>0.61</td>
<td>–0.66</td>
<td>0.06</td>
</tr>
<tr>
<td>8.28</td>
<td>8.42</td>
<td>0.66</td>
<td>–0.62</td>
<td>0.04</td>
</tr>
<tr>
<td>9.78</td>
<td>9.96</td>
<td>0.73</td>
<td>–0.58</td>
<td>0.04</td>
</tr>
<tr>
<td>11.29</td>
<td>11.50</td>
<td>0.87</td>
<td>–0.56</td>
<td>0.03</td>
</tr>
<tr>
<td>12.79</td>
<td>13.06</td>
<td>1.01</td>
<td>–0.55</td>
<td>0.02</td>
</tr>
<tr>
<td>14.30</td>
<td>14.65</td>
<td>1.29</td>
<td>–0.54</td>
<td>0.02</td>
</tr>
<tr>
<td>15.80</td>
<td>16.20</td>
<td>1.46</td>
<td>–0.53</td>
<td>0.02</td>
</tr>
<tr>
<td>17.31</td>
<td>17.90</td>
<td>1.73</td>
<td>–0.52</td>
<td>0.01</td>
</tr>
<tr>
<td>18.81</td>
<td>19.43</td>
<td>1.78</td>
<td>–0.52</td>
<td>0.01</td>
</tr>
<tr>
<td>20.32</td>
<td>20.96</td>
<td>1.71</td>
<td>–0.52</td>
<td>0.01</td>
</tr>
<tr>
<td>21.82</td>
<td>21.21</td>
<td>1.49</td>
<td>–0.51</td>
<td>0.01</td>
</tr>
<tr>
<td>23.33</td>
<td>22.87</td>
<td>1.06</td>
<td>–0.51</td>
<td>0.01</td>
</tr>
<tr>
<td>Actual F0 and Slope: 273.19 Hz and –6.11 dB/octave</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.75</td>
<td>0.97</td>
<td>0.58</td>
<td>–1.04</td>
<td>0.16</td>
</tr>
<tr>
<td>2.26</td>
<td>2.37</td>
<td>0.66</td>
<td>–0.87</td>
<td>0.12</td>
</tr>
<tr>
<td>3.76</td>
<td>3.87</td>
<td>0.65</td>
<td>–0.74</td>
<td>0.09</td>
</tr>
<tr>
<td>5.27</td>
<td>5.39</td>
<td>0.66</td>
<td>–0.65</td>
<td>0.07</td>
</tr>
<tr>
<td>6.77</td>
<td>6.91</td>
<td>0.71</td>
<td>–0.59</td>
<td>0.05</td>
</tr>
<tr>
<td>8.28</td>
<td>8.43</td>
<td>0.76</td>
<td>–0.54</td>
<td>0.04</td>
</tr>
<tr>
<td>9.78</td>
<td>9.96</td>
<td>0.85</td>
<td>–0.51</td>
<td>0.03</td>
</tr>
<tr>
<td>11.29</td>
<td>11.51</td>
<td>0.90</td>
<td>–0.48</td>
<td>0.03</td>
</tr>
<tr>
<td>12.79</td>
<td>13.06</td>
<td>1.11</td>
<td>–0.47</td>
<td>0.02</td>
</tr>
<tr>
<td>14.30</td>
<td>14.62</td>
<td>1.38</td>
<td>–0.45</td>
<td>0.02</td>
</tr>
<tr>
<td>15.80</td>
<td>16.30</td>
<td>1.70</td>
<td>–0.44</td>
<td>0.02</td>
</tr>
<tr>
<td>17.31</td>
<td>18.06</td>
<td>2.06</td>
<td>–0.44</td>
<td>0.02</td>
</tr>
<tr>
<td>18.81</td>
<td>19.52</td>
<td>2.15</td>
<td>–0.43</td>
<td>0.01</td>
</tr>
<tr>
<td>20.32</td>
<td>20.74</td>
<td>2.05</td>
<td>–0.43</td>
<td>0.01</td>
</tr>
<tr>
<td>21.82</td>
<td>21.42</td>
<td>1.81</td>
<td>–0.43</td>
<td>0.01</td>
</tr>
<tr>
<td>23.33</td>
<td>21.93</td>
<td>1.65</td>
<td>–0.43</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Table 5.8.4 Means and standard deviations (s.d.) of the measured HNR and log pACC as a function of the sixteen model HNRs generated at 229.72 Hz and 273.19 Hz. The actual harmonic slope is –6.11 dB/slope.
<table>
<thead>
<tr>
<th>Actual HNR (dB)</th>
<th>Measured HNR (dB)</th>
<th>Log pACC (dB)</th>
<th>F0 (Hz)</th>
<th>Slope (dB/octave)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>s.d.</td>
<td>mean</td>
<td>s.d.</td>
</tr>
<tr>
<td>Actual F0 and Slope: 72.88 Hz and –12.23 dB/octave</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.75</td>
<td>0.84</td>
<td>0.35</td>
<td>–2.09</td>
<td>0.16</td>
</tr>
<tr>
<td>2.26</td>
<td>2.24</td>
<td>0.44</td>
<td>–2.02</td>
<td>0.14</td>
</tr>
<tr>
<td>3.76</td>
<td>3.77</td>
<td>0.35</td>
<td>–1.97</td>
<td>0.12</td>
</tr>
<tr>
<td>5.27</td>
<td>5.28</td>
<td>0.30</td>
<td>–1.94</td>
<td>0.11</td>
</tr>
<tr>
<td>6.77</td>
<td>6.78</td>
<td>0.28</td>
<td>–1.91</td>
<td>0.10</td>
</tr>
<tr>
<td>8.28</td>
<td>8.28</td>
<td>0.26</td>
<td>–1.89</td>
<td>0.09</td>
</tr>
<tr>
<td>9.78</td>
<td>9.79</td>
<td>0.30</td>
<td>–1.88</td>
<td>0.08</td>
</tr>
<tr>
<td>11.29</td>
<td>11.30</td>
<td>0.39</td>
<td>–1.87</td>
<td>0.08</td>
</tr>
<tr>
<td>12.79</td>
<td>12.79</td>
<td>0.56</td>
<td>–1.86</td>
<td>0.08</td>
</tr>
<tr>
<td>14.30</td>
<td>14.33</td>
<td>0.70</td>
<td>–1.86</td>
<td>0.07</td>
</tr>
<tr>
<td>15.80</td>
<td>15.80</td>
<td>0.88</td>
<td>–1.86</td>
<td>0.07</td>
</tr>
<tr>
<td>17.31</td>
<td>17.31</td>
<td>1.12</td>
<td>–1.86</td>
<td>0.07</td>
</tr>
<tr>
<td>18.81</td>
<td>18.83</td>
<td>1.67</td>
<td>–1.85</td>
<td>0.07</td>
</tr>
<tr>
<td>20.32</td>
<td>20.28</td>
<td>1.79</td>
<td>–1.85</td>
<td>0.07</td>
</tr>
<tr>
<td>21.82</td>
<td>21.37</td>
<td>1.82</td>
<td>–1.85</td>
<td>0.07</td>
</tr>
<tr>
<td>23.33</td>
<td>22.09</td>
<td>1.51</td>
<td>–1.85</td>
<td>0.07</td>
</tr>
<tr>
<td>Actual F0 and Slope: 128.93 Hz and –12.23 dB/octave</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.75</td>
<td>0.82</td>
<td>0.30</td>
<td>–1.24</td>
<td>0.13</td>
</tr>
<tr>
<td>2.26</td>
<td>2.26</td>
<td>0.38</td>
<td>–1.16</td>
<td>0.10</td>
</tr>
<tr>
<td>3.76</td>
<td>3.77</td>
<td>0.34</td>
<td>–1.10</td>
<td>0.08</td>
</tr>
<tr>
<td>5.27</td>
<td>5.27</td>
<td>0.28</td>
<td>–1.05</td>
<td>0.07</td>
</tr>
<tr>
<td>6.77</td>
<td>6.79</td>
<td>0.25</td>
<td>–1.02</td>
<td>0.06</td>
</tr>
<tr>
<td>8.28</td>
<td>8.30</td>
<td>0.28</td>
<td>–1.00</td>
<td>0.05</td>
</tr>
<tr>
<td>9.78</td>
<td>9.79</td>
<td>0.28</td>
<td>–0.99</td>
<td>0.04</td>
</tr>
<tr>
<td>11.29</td>
<td>11.30</td>
<td>0.32</td>
<td>–0.98</td>
<td>0.04</td>
</tr>
<tr>
<td>12.79</td>
<td>12.79</td>
<td>0.40</td>
<td>–0.97</td>
<td>0.03</td>
</tr>
<tr>
<td>14.30</td>
<td>14.29</td>
<td>0.55</td>
<td>–0.96</td>
<td>0.03</td>
</tr>
<tr>
<td>15.80</td>
<td>15.84</td>
<td>0.75</td>
<td>–0.96</td>
<td>0.03</td>
</tr>
<tr>
<td>17.31</td>
<td>17.39</td>
<td>0.96</td>
<td>–0.96</td>
<td>0.02</td>
</tr>
<tr>
<td>18.81</td>
<td>19.01</td>
<td>1.20</td>
<td>–0.95</td>
<td>0.02</td>
</tr>
<tr>
<td>20.32</td>
<td>20.65</td>
<td>1.61</td>
<td>–0.95</td>
<td>0.02</td>
</tr>
<tr>
<td>21.82</td>
<td>21.91</td>
<td>1.62</td>
<td>–0.95</td>
<td>0.02</td>
</tr>
<tr>
<td>23.33</td>
<td>22.51</td>
<td>1.40</td>
<td>–0.95</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Table 5.8.5 Means and standard deviations (s.d.) of the measured HNR and log pACC as a function of the sixteen model HNRs generated at 72.88 Hz and 128.93 Hz. The actual harmonic slope is –12.23 dB/slope.
<table>
<thead>
<tr>
<th>Actual HNR (dB)</th>
<th>Measured HNR (dB)</th>
<th>Log pACC (dB)</th>
<th>F0 (Hz)</th>
<th>Slope (dB/octave)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>s.d.</td>
<td>mean</td>
<td>s.d.</td>
</tr>
<tr>
<td>Actual F0 and Slope: 229.72 Hz and –12.23 dB/octave</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.75</td>
<td>0.80</td>
<td>0.27</td>
<td>–0.97</td>
<td>0.13</td>
</tr>
<tr>
<td>2.26</td>
<td>2.29</td>
<td>0.24</td>
<td>–0.84</td>
<td>0.10</td>
</tr>
<tr>
<td>3.76</td>
<td>3.79</td>
<td>0.22</td>
<td>–0.75</td>
<td>0.08</td>
</tr>
<tr>
<td>5.27</td>
<td>5.29</td>
<td>0.20</td>
<td>–0.68</td>
<td>0.06</td>
</tr>
<tr>
<td>6.77</td>
<td>6.79</td>
<td>0.20</td>
<td>–0.63</td>
<td>0.05</td>
</tr>
<tr>
<td>8.28</td>
<td>8.30</td>
<td>0.20</td>
<td>–0.59</td>
<td>0.04</td>
</tr>
<tr>
<td>9.78</td>
<td>9.80</td>
<td>0.24</td>
<td>–0.57</td>
<td>0.03</td>
</tr>
<tr>
<td>11.29</td>
<td>11.31</td>
<td>0.27</td>
<td>–0.55</td>
<td>0.03</td>
</tr>
<tr>
<td>12.79</td>
<td>12.83</td>
<td>0.34</td>
<td>–0.54</td>
<td>0.02</td>
</tr>
<tr>
<td>14.30</td>
<td>14.36</td>
<td>0.44</td>
<td>–0.53</td>
<td>0.02</td>
</tr>
<tr>
<td>15.80</td>
<td>15.92</td>
<td>0.59</td>
<td>–0.52</td>
<td>0.02</td>
</tr>
<tr>
<td>17.31</td>
<td>17.51</td>
<td>0.77</td>
<td>–0.52</td>
<td>0.01</td>
</tr>
<tr>
<td>18.81</td>
<td>19.16</td>
<td>1.01</td>
<td>–0.52</td>
<td>0.01</td>
</tr>
<tr>
<td>20.32</td>
<td>20.81</td>
<td>1.24</td>
<td>–0.51</td>
<td>0.01</td>
</tr>
<tr>
<td>21.82</td>
<td>22.29</td>
<td>1.26</td>
<td>–0.51</td>
<td>0.01</td>
</tr>
<tr>
<td>23.33</td>
<td>22.90</td>
<td>1.02</td>
<td>–0.51</td>
<td>0.01</td>
</tr>
<tr>
<td>Actual F0 and Slope: 273.19 Hz and –12.23 dB/octave</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.75</td>
<td>0.82</td>
<td>0.30</td>
<td>–0.93</td>
<td>0.13</td>
</tr>
<tr>
<td>2.26</td>
<td>2.29</td>
<td>0.36</td>
<td>–0.78</td>
<td>0.10</td>
</tr>
<tr>
<td>3.76</td>
<td>3.79</td>
<td>0.33</td>
<td>–0.68</td>
<td>0.08</td>
</tr>
<tr>
<td>5.27</td>
<td>5.29</td>
<td>0.28</td>
<td>–0.61</td>
<td>0.06</td>
</tr>
<tr>
<td>6.77</td>
<td>6.79</td>
<td>0.28</td>
<td>–0.55</td>
<td>0.05</td>
</tr>
<tr>
<td>8.28</td>
<td>8.29</td>
<td>0.28</td>
<td>–0.52</td>
<td>0.04</td>
</tr>
<tr>
<td>9.78</td>
<td>9.82</td>
<td>0.35</td>
<td>–0.49</td>
<td>0.03</td>
</tr>
<tr>
<td>11.29</td>
<td>11.36</td>
<td>0.45</td>
<td>–0.47</td>
<td>0.03</td>
</tr>
<tr>
<td>12.79</td>
<td>12.92</td>
<td>0.56</td>
<td>–0.46</td>
<td>0.02</td>
</tr>
<tr>
<td>14.30</td>
<td>14.49</td>
<td>0.69</td>
<td>–0.45</td>
<td>0.02</td>
</tr>
<tr>
<td>15.80</td>
<td>16.09</td>
<td>0.87</td>
<td>–0.44</td>
<td>0.01</td>
</tr>
<tr>
<td>17.31</td>
<td>17.66</td>
<td>1.17</td>
<td>–0.44</td>
<td>0.01</td>
</tr>
<tr>
<td>18.81</td>
<td>19.37</td>
<td>1.61</td>
<td>–0.43</td>
<td>0.01</td>
</tr>
<tr>
<td>20.32</td>
<td>20.97</td>
<td>1.76</td>
<td>–0.43</td>
<td>0.01</td>
</tr>
<tr>
<td>21.82</td>
<td>21.95</td>
<td>1.51</td>
<td>–0.43</td>
<td>0.01</td>
</tr>
<tr>
<td>23.33</td>
<td>22.53</td>
<td>1.20</td>
<td>–0.43</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Table 5.8.6 Means and standard deviations (s.d.) of the measured HNR and log pACC as a function of the sixteen model HNRs generated at 229.72 Hz and 273.19 Hz. The actual harmonic slope is –12.23 dB/slope.
5.4.2 Evaluation of the ASHE algorithm using the glottal wave

The ASHE procedure is also evaluated by the use of the differentiated glottal wave with varying levels of additive noise on simultaneous jitter and shimmer pedestals. Sixteen HNRs are generated at two fundamental frequencies (128.92, 229.72 Hz) with two levels of simultaneous jitter and shimmer: 1.0%/0.56 dB and 2.5%/1.32 dB. The data are presented in Tables 5.10.1-5.10.2 below. The results indicate that the ASHE technique performs more poorly than when it is applied to the model signals used in the synthesis phase. Nevertheless, the Pearson correlation coefficients in Table 5.11 reach values nearly equivalent to those of the model signals in Table 5.9.

With simultaneous modulation noise typical of normal voice (1.0%/0.56 dB), the actual HNRs above about 18 dB tend to be assessed at HNRs manifestly higher than their true values. This means that substantial aspiration noise (HNR < 18 dB) must be present in the speech signal in order for it to be detected as such in a consistent and reliable manner. Hence the HNR threshold for the phonological significance of aspiration noise must be 18 dB or less. Note that both Kojima et al. (1980) and Klingholz (1987) found the lower HNR limit of normal voice to be about 15 dB (Section 3.2.4). As a

<table>
<thead>
<tr>
<th>Slope (dB/octave)</th>
<th>72.88 Hz</th>
<th>128.92 Hz</th>
<th>229.72 Hz</th>
<th>273.18 Hz</th>
</tr>
</thead>
<tbody>
<tr>
<td>–3.06</td>
<td>0.990</td>
<td>&lt;0.001</td>
<td>0.995</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>–6.11</td>
<td>0.999</td>
<td>&lt;0.001</td>
<td>0.999</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>–12.23</td>
<td>0.999</td>
<td>&lt;0.001</td>
<td>0.999</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

**Table 5.9** Pearson product moment correlation coefficients between the actual and mean HNRs of the model signals at four F0s. All correlation coefficients are significant at less than the 0.001 level (N=16).
consequence, the upper threshold range for aspiration noise is probably on the order of 15-18 dB. Rather unexpectedly, the presence of large amounts of modulation noise (2.5%/1.32 dB) leads to better HNR estimates in the same range. The larger Pearson correlation coefficients of the harsh voice waveforms substantiate this trend. Although the overall performance is somewhat diminished when the test signals are no longer the ones used during the synthesis phase, the HNR estimates of the ASHE technique are thus further demonstrated to be insensitive to fairly large amounts of modulation noise.
<table>
<thead>
<tr>
<th>Actual HNR (dB)</th>
<th>Measured HNR (dB)</th>
<th>Log pACC (dB)</th>
<th>F0 (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>s.d.</td>
<td>mean</td>
</tr>
<tr>
<td>Actual F0, jitter, and shimmer: 128.93 Hz, 1.00%, and 0.56 dB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.75</td>
<td>0.78</td>
<td>0.23</td>
<td>-1.74</td>
</tr>
<tr>
<td>2.26</td>
<td>0.85</td>
<td>0.40</td>
<td>-1.54</td>
</tr>
<tr>
<td>3.76</td>
<td>1.71</td>
<td>0.91</td>
<td>-1.40</td>
</tr>
<tr>
<td>5.27</td>
<td>3.28</td>
<td>0.88</td>
<td>-1.30</td>
</tr>
<tr>
<td>6.77</td>
<td>4.90</td>
<td>0.87</td>
<td>-1.23</td>
</tr>
<tr>
<td>8.28</td>
<td>6.60</td>
<td>0.85</td>
<td>-1.17</td>
</tr>
<tr>
<td>9.78</td>
<td>8.36</td>
<td>0.82</td>
<td>-1.14</td>
</tr>
<tr>
<td>11.29</td>
<td>10.22</td>
<td>0.86</td>
<td>-1.11</td>
</tr>
<tr>
<td>12.79</td>
<td>12.27</td>
<td>1.07</td>
<td>-1.09</td>
</tr>
<tr>
<td>14.30</td>
<td>14.68</td>
<td>1.48</td>
<td>-1.08</td>
</tr>
<tr>
<td>15.80</td>
<td>17.95</td>
<td>2.33</td>
<td>-1.07</td>
</tr>
<tr>
<td>17.31</td>
<td>21.41</td>
<td>2.03</td>
<td>-1.06</td>
</tr>
<tr>
<td>18.81</td>
<td>23.11</td>
<td>0.70</td>
<td>-1.05</td>
</tr>
<tr>
<td>20.32</td>
<td>23.32</td>
<td>0.10</td>
<td>-1.05</td>
</tr>
<tr>
<td>21.82</td>
<td>23.33</td>
<td>0.00</td>
<td>-1.05</td>
</tr>
<tr>
<td>23.33</td>
<td>23.33</td>
<td>0.00</td>
<td>-1.05</td>
</tr>
<tr>
<td>Actual F0, jitter, and shimmer: 128.93 Hz, 2.50%, and 1.32 dB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.75</td>
<td>0.94</td>
<td>0.99</td>
<td>-2.18</td>
</tr>
<tr>
<td>2.26</td>
<td>0.86</td>
<td>0.43</td>
<td>-1.99</td>
</tr>
<tr>
<td>3.76</td>
<td>1.63</td>
<td>0.93</td>
<td>-1.84</td>
</tr>
<tr>
<td>5.27</td>
<td>3.12</td>
<td>0.98</td>
<td>-1.74</td>
</tr>
<tr>
<td>6.77</td>
<td>4.67</td>
<td>0.98</td>
<td>-1.66</td>
</tr>
<tr>
<td>8.28</td>
<td>6.34</td>
<td>1.01</td>
<td>-1.61</td>
</tr>
<tr>
<td>9.78</td>
<td>8.06</td>
<td>1.08</td>
<td>-1.57</td>
</tr>
<tr>
<td>11.29</td>
<td>9.82</td>
<td>1.15</td>
<td>-1.54</td>
</tr>
<tr>
<td>12.79</td>
<td>11.64</td>
<td>1.29</td>
<td>-1.52</td>
</tr>
<tr>
<td>14.30</td>
<td>13.75</td>
<td>1.62</td>
<td>-1.50</td>
</tr>
<tr>
<td>15.80</td>
<td>16.25</td>
<td>2.32</td>
<td>-1.49</td>
</tr>
<tr>
<td>17.31</td>
<td>19.13</td>
<td>2.84</td>
<td>-1.49</td>
</tr>
<tr>
<td>18.81</td>
<td>21.40</td>
<td>2.33</td>
<td>-1.48</td>
</tr>
<tr>
<td>20.32</td>
<td>21.40</td>
<td>2.33</td>
<td>-1.48</td>
</tr>
<tr>
<td>21.82</td>
<td>23.06</td>
<td>0.86</td>
<td>-1.47</td>
</tr>
<tr>
<td>23.33</td>
<td>23.23</td>
<td>0.51</td>
<td>-1.47</td>
</tr>
</tbody>
</table>

**Table 5.10.1** Means and standard deviations (s.d.) of the measured HNR and log pACC of the differentiated glottal wave as a function of sixteen HNRs generated at 128.93 Hz. The simultaneous jitter and shimmer pedestals are respectively 1.0%/0.56 dB (top) and 2.50%/1.32 dB (bottom).
<table>
<thead>
<tr>
<th>Actual HNR (dB)</th>
<th>Measured HNR (dB)</th>
<th>Log pACC (dB)</th>
<th>F0 (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>s.d.</td>
<td>mean</td>
</tr>
<tr>
<td>Actual F0, jitter, and shimmer: 229.72 Hz, 1.00%, and 0.56 dB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.75</td>
<td>0.78</td>
<td>0.21</td>
<td>−1.67</td>
</tr>
<tr>
<td>2.26</td>
<td>0.87</td>
<td>0.47</td>
<td>−1.38</td>
</tr>
<tr>
<td>3.76</td>
<td>1.88</td>
<td>0.92</td>
<td>−1.16</td>
</tr>
<tr>
<td>5.27</td>
<td>3.49</td>
<td>0.91</td>
<td>−1.01</td>
</tr>
<tr>
<td>6.77</td>
<td>5.17</td>
<td>0.92</td>
<td>−0.89</td>
</tr>
<tr>
<td>8.28</td>
<td>6.92</td>
<td>0.96</td>
<td>−0.81</td>
</tr>
<tr>
<td>9.78</td>
<td>8.82</td>
<td>1.17</td>
<td>−0.75</td>
</tr>
<tr>
<td>11.29</td>
<td>10.85</td>
<td>1.45</td>
<td>−0.71</td>
</tr>
<tr>
<td>12.79</td>
<td>13.48</td>
<td>2.09</td>
<td>−0.68</td>
</tr>
<tr>
<td>14.30</td>
<td>16.88</td>
<td>2.89</td>
<td>−0.66</td>
</tr>
<tr>
<td>15.80</td>
<td>20.69</td>
<td>2.73</td>
<td>−0.65</td>
</tr>
<tr>
<td>17.31</td>
<td>22.67</td>
<td>1.44</td>
<td>−0.64</td>
</tr>
<tr>
<td>18.81</td>
<td>23.27</td>
<td>0.38</td>
<td>−0.63</td>
</tr>
<tr>
<td>20.32</td>
<td>23.33</td>
<td>0.07</td>
<td>−0.62</td>
</tr>
<tr>
<td>21.82</td>
<td>23.33</td>
<td>0.07</td>
<td>−0.62</td>
</tr>
<tr>
<td>23.33</td>
<td>23.33</td>
<td>0.00</td>
<td>−0.62</td>
</tr>
<tr>
<td>Actual F0, jitter, and shimmer: 229.72 Hz, 2.50%, and 1.32 dB</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.75</td>
<td>0.77</td>
<td>0.16</td>
<td>−2.14</td>
</tr>
<tr>
<td>2.26</td>
<td>0.83</td>
<td>0.37</td>
<td>−1.86</td>
</tr>
<tr>
<td>3.76</td>
<td>1.63</td>
<td>0.89</td>
<td>−1.65</td>
</tr>
<tr>
<td>5.27</td>
<td>3.09</td>
<td>1.01</td>
<td>−1.50</td>
</tr>
<tr>
<td>6.77</td>
<td>4.64</td>
<td>0.99</td>
<td>−1.39</td>
</tr>
<tr>
<td>8.28</td>
<td>6.23</td>
<td>1.05</td>
<td>−1.31</td>
</tr>
<tr>
<td>9.78</td>
<td>7.89</td>
<td>1.20</td>
<td>−1.25</td>
</tr>
<tr>
<td>11.29</td>
<td>9.64</td>
<td>1.34</td>
<td>−1.21</td>
</tr>
<tr>
<td>12.79</td>
<td>11.42</td>
<td>1.58</td>
<td>−1.18</td>
</tr>
<tr>
<td>14.30</td>
<td>13.50</td>
<td>2.16</td>
<td>−1.16</td>
</tr>
<tr>
<td>15.80</td>
<td>15.93</td>
<td>3.03</td>
<td>−1.14</td>
</tr>
<tr>
<td>17.31</td>
<td>18.41</td>
<td>3.53</td>
<td>−1.13</td>
</tr>
<tr>
<td>18.81</td>
<td>20.21</td>
<td>3.30</td>
<td>−1.13</td>
</tr>
<tr>
<td>20.32</td>
<td>21.33</td>
<td>2.90</td>
<td>−1.12</td>
</tr>
<tr>
<td>21.82</td>
<td>21.94</td>
<td>2.55</td>
<td>−1.12</td>
</tr>
<tr>
<td>23.33</td>
<td>22.21</td>
<td>2.27</td>
<td>−1.11</td>
</tr>
</tbody>
</table>

Table 5.10.2 Means and standard deviations (s.d.) of the measured HNR and log pACC of the differentiated glottal wave as a function of sixteen HNRs generated at 229.72 Hz. The simultaneous jitter and shimmer pedestals are respectively 1.0%/0.56 dB (top) and 2.50%/1.32 dB (bottom).
5.5 Performance of the spectral tilt measures

A serious disadvantage of the mid-band slopes H1–F1, H1–F2, H1–F3 is that large amounts of aspiration noise can introduce a great deal of variability in the measurement of the formant levels. Since the noise is explicitly modeled by the analysis-by-synthesis procedure, the harmonic slope should be less susceptible to this kind of error. The means and standard deviations of the harmonic slope over the full range of the synthesized HNRs are given in Tables 5.8.1-5.8.6. The results show that the actual harmonic slopes (−3.05, −6.11, and −12.23 dB/octave) and the means of the measured harmonic slopes are very similar. Even at the lowest HNR of 0.75 dB, the standard deviations represent a fairly small degree of scatter relative to the mean, the largest coefficient of variation (the standard deviation/mean × 100) being 15.84% at 72.88 Hz and −3.05 dB/octave. Hence the ASHE technique provides a good estimate of mid to high frequency spectral tilt even when additive noise levels are significant.
To provide an indication of the behavior of the wide-band harmonic slope in normal speech, the quadrangular vowel system [a æ i u] is examined. Hillenbrand et al. (1995) recorded a set of American English vowels at a sampling rate of 16 kHz. The vowels were produced in the environment /hVd/. The first twenty adult male and female tokens of the four vowels were taken from the database following its original order. The acoustic parameters are evaluated at the vowel frame with maximum energy relative to to the other vowel frames. For men and women speakers, the means and standard deviations of the acoustic parameters of the four natural vowels are shown in Table 5.12. Note the well-known tendency for the high vowels to have significantly higher mean F0s than the low vowels.

<table>
<thead>
<tr>
<th>Vowels</th>
<th>F0 (Hz)</th>
<th>Slope (dB/octave)</th>
<th>H1–H2 (dB)</th>
<th>Log pACC (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>mean</td>
<td>s.d.</td>
<td>mean</td>
<td>s.d.</td>
</tr>
<tr>
<td>Natural Men</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>124.97</td>
<td>19.40</td>
<td>−3.49</td>
<td>0.18</td>
</tr>
<tr>
<td>æ</td>
<td>124.43</td>
<td>21.03</td>
<td>−3.60</td>
<td>0.18</td>
</tr>
<tr>
<td>i</td>
<td>137.09</td>
<td>24.40</td>
<td>−4.19</td>
<td>0.66</td>
</tr>
<tr>
<td>u</td>
<td>138.33</td>
<td>23.45</td>
<td>−4.42</td>
<td>0.18</td>
</tr>
<tr>
<td>Synthetic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>125.26</td>
<td>0.00</td>
<td>−3.53</td>
<td>0.18</td>
</tr>
<tr>
<td>æ</td>
<td>125.26</td>
<td>0.00</td>
<td>−4.00</td>
<td>0.01</td>
</tr>
<tr>
<td>i</td>
<td>136.59</td>
<td>0.00</td>
<td>−5.22</td>
<td>0.02</td>
</tr>
<tr>
<td>u</td>
<td>138.58</td>
<td>0.00</td>
<td>−4.66</td>
<td>0.01</td>
</tr>
<tr>
<td>Natural Women</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>215.29</td>
<td>19.06</td>
<td>−3.27</td>
<td>0.18</td>
</tr>
<tr>
<td>æ</td>
<td>216.59</td>
<td>21.12</td>
<td>−3.56</td>
<td>0.20</td>
</tr>
<tr>
<td>i</td>
<td>230.97</td>
<td>21.35</td>
<td>−4.82</td>
<td>1.66</td>
</tr>
<tr>
<td>u</td>
<td>236.42</td>
<td>17.69</td>
<td>−4.49</td>
<td>0.17</td>
</tr>
</tbody>
</table>

Table 5.12 Means and standard deviations (s.d.) of the acoustic parameters of natural and synthetic American English vowels (Hillenbrand et al., 1995; see text for the generation of the synthetic vowels).
To determine how the formant structure of vowels affects spectral tilt, the differentiated glottal waveforms are fed into a cascade of five digital resonators (see Equation 4.1 and Klatt, 1980 for the method). The formant frequencies and 3 dB bandwidths are fixed at the men’s values furnished by Hillenbrand et al. (1995) and Klatt (1980). The radiation characteristic is not modeled since a +6 dB/octave boost is effectively incorporated into the glottal source. The synthetic vowels [a æ i u] are generated at the men’s natural F0s given in Table 5.12 and the four waveforms are illustrated in Figure 5.4. Like the natural vowels, the means and standard deviations of the synthetic vowels are presented in Table 5.12. Comparison of the natural and synthetic vowel measures shows fairly good parallels between them, with the exception of the much lower H1–H2 value of the natural vowels. The most striking result is the flatter wide-band harmonic slope of the low vowels, particularly [a]. Clearly the formant structure alone, particularly a high F1, produces this effect because the synthetic vowels are generated with similarly shaped glottal waveforms differing only in F0. Recall that meaningful comparison of spectral tilt presupposes vowels with like formant structures.
In Chapter 4, a parameter called the peak energy factor PE was proposed as a method of approximating the quality factor Q. Defined in Equation 4.10, the PE has three properties in common with Q: it is dimensionless, invariant with respect to multiplicative frequency shifts, and serves as an indicator of the amplification at resonance. As was observed in Section 4.5, linear prediction of the speech wave is the current method for determining formant bandwidths, and thus formant Qs.

The goal of this section is to compare the results of the peak energy method with those obtained through linear prediction (LP) by using signals with known resonant frequencies and bandwidths. Two 0.5 s source signals are generated at a sampling rate of 44.1 kHz, one a 130 Hz sawtooth, the other a white noise waveform. Both signals are fed into the second-order digital resonator described by Equation 4.1, with the damped frequency \( f_d \) being set equal to the resonant frequency \( f_r \) of 500 Hz. Eleven actual 3 dB
bandwidths ($B_{3dB}$) are used, starting with a 3 dB bandwidth of 500 Hz, then successively dividing it by the square root of 2, for a final 3 dB bandwidth of 15.63 Hz. The 3 dB bandwidths are likewise expressed as $Q_{3dB}$ and $20\log_{10}Q_{3dB}$ (see Tables 5.13.1-5.13.2).

These synthetic first formant waveforms are analyzed by the peak energy and linear prediction procedures. To determine the first formant frequency and bandwidth according to the linear predictive method, the MATLAB statement $a = \text{lpc}(x,p)$ is employed, where $x$ is the signal vector corresponding to the 39.37 ms frame of the present method, $a$ the vector of linear prediction coefficients, and $p$ the predictor order fixed at 12, a typical value in speech analysis. The time step for both methods is 2.46 ms, resulting in a total of 186 frames. The roots of the predictor polynomial are then computed by the MATLAB statement $z = \text{roots}(a)$. Possible first formant frequencies ($f_1$) and bandwidths ($b_1$) are determined through the respective MATLAB statements $f_1 = \text{imag}(\log(z)) \ast (\text{fs}/(2\ast\pi))$ and $b_1 = \text{real}(\log(z)) \ast (-\text{fs}/\pi)$, where $\text{fs}$ is the 44.1 kHz sampling frequency. Finally, the $b_1$ vector of each frame is scanned to find the minimum value, and the associated $f_1$ is considered to be the first formant frequency (F1). Note that in the peak energy method, F1 is also found by searching the formant range for the $B_1$ minimum, which for all intents and purposes is the same as the maximum value of $PE_1$ (Section 4.5).
# 130 Hz sawtooth source and a 500 Hz resonant frequency

## Peak Energy Method (PE)

<table>
<thead>
<tr>
<th>Actual B&lt;sub&gt;3dB&lt;/sub&gt;</th>
<th>Actual Q&lt;sub&gt;3dB&lt;/sub&gt;</th>
<th>Actual 20log&lt;sub&gt;10&lt;/sub&gt;Q&lt;sub&gt;3dB&lt;/sub&gt;</th>
<th>Frequency Mean (Hz)</th>
<th>Frequency s.d. (Hz)</th>
<th>PE mean (dB)</th>
<th>PE s.d. (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>500.00</td>
<td>1.000</td>
<td>0.00</td>
<td>285.44</td>
<td>50.52</td>
<td>4.9688</td>
<td>0.1375</td>
</tr>
<tr>
<td>353.55</td>
<td>1.414</td>
<td>3.01</td>
<td>391.97</td>
<td>0.00</td>
<td>5.2322</td>
<td>0.1262</td>
</tr>
<tr>
<td>250.00</td>
<td>2.000</td>
<td>6.02</td>
<td>391.97</td>
<td>0.00</td>
<td>5.5005</td>
<td>0.1306</td>
</tr>
<tr>
<td>176.66</td>
<td>2.828</td>
<td>9.03</td>
<td>523.21</td>
<td>0.00</td>
<td>5.9672</td>
<td>0.1071</td>
</tr>
<tr>
<td>125.00</td>
<td>4.000</td>
<td>12.04</td>
<td>523.21</td>
<td>0.00</td>
<td>6.6048</td>
<td>0.1175</td>
</tr>
<tr>
<td>88.39</td>
<td>5.657</td>
<td>15.05</td>
<td>523.21</td>
<td>0.00</td>
<td>7.0950</td>
<td>0.1285</td>
</tr>
<tr>
<td>62.50</td>
<td>8.000</td>
<td>18.06</td>
<td>523.21</td>
<td>0.00</td>
<td>7.4188</td>
<td>0.1387</td>
</tr>
<tr>
<td>44.19</td>
<td>11.314</td>
<td>21.07</td>
<td>523.05</td>
<td>1.09</td>
<td>7.6105</td>
<td>0.1463</td>
</tr>
<tr>
<td>31.25</td>
<td>16.000</td>
<td>24.08</td>
<td>522.93</td>
<td>1.43</td>
<td>7.7183</td>
<td>0.1417</td>
</tr>
<tr>
<td>22.09</td>
<td>22.627</td>
<td>27.09</td>
<td>522.85</td>
<td>1.61</td>
<td>7.7776</td>
<td>0.1358</td>
</tr>
<tr>
<td>15.63</td>
<td>32.000</td>
<td>30.10</td>
<td>522.81</td>
<td>1.70</td>
<td>7.8094</td>
<td>0.1337</td>
</tr>
</tbody>
</table>

## Linear Prediction Method (LP)

<table>
<thead>
<tr>
<th>Actual B&lt;sub&gt;3dB&lt;/sub&gt;</th>
<th>Actual Q&lt;sub&gt;3dB&lt;/sub&gt;</th>
<th>Actual 20log&lt;sub&gt;10&lt;/sub&gt;Q&lt;sub&gt;3dB&lt;/sub&gt;</th>
<th>Frequency Mean (Hz)</th>
<th>Frequency s.d. (Hz)</th>
<th>B&lt;sub&gt;3dB&lt;/sub&gt; mean (Hz)</th>
<th>B&lt;sub&gt;3dB&lt;/sub&gt; s.d. (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>500.00</td>
<td>1.000</td>
<td>0.00</td>
<td>241.55</td>
<td>70.06</td>
<td>298.83</td>
<td>83.54</td>
</tr>
<tr>
<td>353.55</td>
<td>1.414</td>
<td>3.01</td>
<td>297.67</td>
<td>44.13</td>
<td>264.19</td>
<td>66.91</td>
</tr>
<tr>
<td>250.00</td>
<td>2.000</td>
<td>6.02</td>
<td>326.80</td>
<td>40.90</td>
<td>228.85</td>
<td>50.09</td>
</tr>
<tr>
<td>176.66</td>
<td>2.828</td>
<td>9.03</td>
<td>356.94</td>
<td>34.83</td>
<td>194.30</td>
<td>42.43</td>
</tr>
<tr>
<td>125.00</td>
<td>4.000</td>
<td>12.04</td>
<td>386.08</td>
<td>24.34</td>
<td>162.13</td>
<td>35.58</td>
</tr>
<tr>
<td>88.39</td>
<td>5.657</td>
<td>15.05</td>
<td>415.99</td>
<td>17.31</td>
<td>130.24</td>
<td>27.94</td>
</tr>
<tr>
<td>62.50</td>
<td>8.000</td>
<td>18.06</td>
<td>442.95</td>
<td>13.06</td>
<td>102.98</td>
<td>22.52</td>
</tr>
<tr>
<td>44.19</td>
<td>11.314</td>
<td>21.07</td>
<td>462.73</td>
<td>10.29</td>
<td>82.64</td>
<td>17.94</td>
</tr>
<tr>
<td>31.25</td>
<td>16.000</td>
<td>24.08</td>
<td>474.95</td>
<td>6.64</td>
<td>70.46</td>
<td>14.96</td>
</tr>
<tr>
<td>22.09</td>
<td>22.627</td>
<td>27.09</td>
<td>482.32</td>
<td>5.80</td>
<td>63.28</td>
<td>13.27</td>
</tr>
<tr>
<td>15.63</td>
<td>32.000</td>
<td>30.10</td>
<td>486.35</td>
<td>5.31</td>
<td>59.58</td>
<td>12.58</td>
</tr>
</tbody>
</table>

**Table 5.13.1** Means and standard deviations (s.d.) of the measured peak energy factor (PE) and the linear prediction 3 dB bandwidth (B<sub>3dB</sub>) as a function of the actual 3 dB bandwidths of a second-order filter tuned to a 500 Hz resonance frequency. The source waveform is a 130 Hz sawtooth.
White noise source and a 500 Hz resonant frequency

<table>
<thead>
<tr>
<th>Actual B&lt;sub&gt;3dB&lt;/sub&gt;</th>
<th>Actual Q&lt;sub&gt;3dB&lt;/sub&gt;</th>
<th>Actual 20log&lt;sub&gt;10&lt;/sub&gt;Q&lt;sub&gt;3dB&lt;/sub&gt;</th>
<th>Frequency Mean (Hz)</th>
<th>Frequency s.d. (Hz)</th>
<th>PE Mean (dB)</th>
<th>PE s.d. (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>500.00</td>
<td>1.000</td>
<td>0.00</td>
<td>464.78</td>
<td>178.68</td>
<td>2.6442</td>
<td>0.6267</td>
</tr>
<tr>
<td>353.55</td>
<td>1.414</td>
<td>3.01</td>
<td>457.59</td>
<td>159.08</td>
<td>2.6527</td>
<td>0.6612</td>
</tr>
<tr>
<td>250.00</td>
<td>2.000</td>
<td>6.02</td>
<td>445.30</td>
<td>137.78</td>
<td>2.6941</td>
<td>0.6964</td>
</tr>
<tr>
<td>176.66</td>
<td>2.828</td>
<td>9.03</td>
<td>453.35</td>
<td>117.32</td>
<td>2.7998</td>
<td>0.7654</td>
</tr>
<tr>
<td>125.00</td>
<td>4.000</td>
<td>12.04</td>
<td>458.32</td>
<td>94.50</td>
<td>3.0478</td>
<td>0.8698</td>
</tr>
<tr>
<td>88.39</td>
<td>5.657</td>
<td>15.05</td>
<td>484.04</td>
<td>61.23</td>
<td>3.4953</td>
<td>0.9267</td>
</tr>
<tr>
<td>62.50</td>
<td>8.000</td>
<td>18.06</td>
<td>492.07</td>
<td>45.54</td>
<td>4.0913</td>
<td>0.9483</td>
</tr>
<tr>
<td>44.19</td>
<td>11.314</td>
<td>21.07</td>
<td>495.36</td>
<td>39.29</td>
<td>4.6856</td>
<td>1.0107</td>
</tr>
<tr>
<td>31.25</td>
<td>16.000</td>
<td>24.08</td>
<td>497.00</td>
<td>35.14</td>
<td>5.2165</td>
<td>1.0841</td>
</tr>
<tr>
<td>22.09</td>
<td>22.627</td>
<td>27.09</td>
<td>498.60</td>
<td>26.09</td>
<td>5.6700</td>
<td>1.0989</td>
</tr>
<tr>
<td>15.63</td>
<td>32.000</td>
<td>30.10</td>
<td>499.83</td>
<td>21.44</td>
<td>6.0506</td>
<td>1.0772</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Actual B&lt;sub&gt;3dB&lt;/sub&gt;</th>
<th>Actual Q&lt;sub&gt;3dB&lt;/sub&gt;</th>
<th>Actual 20log&lt;sub&gt;10&lt;/sub&gt;Q&lt;sub&gt;3dB&lt;/sub&gt;</th>
<th>Frequency Mean (Hz)</th>
<th>Frequency s.d. (Hz)</th>
<th>B&lt;sub&gt;3dB&lt;/sub&gt; Mean (Hz)</th>
<th>B&lt;sub&gt;3dB&lt;/sub&gt; s.d. (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>500.00</td>
<td>1.000</td>
<td>0.00</td>
<td>519.92</td>
<td>36.75</td>
<td>556.21</td>
<td>112.72</td>
</tr>
<tr>
<td>353.55</td>
<td>1.414</td>
<td>3.01</td>
<td>516.83</td>
<td>27.86</td>
<td>402.26</td>
<td>81.60</td>
</tr>
<tr>
<td>250.00</td>
<td>2.000</td>
<td>6.02</td>
<td>514.82</td>
<td>22.63</td>
<td>292.08</td>
<td>62.92</td>
</tr>
<tr>
<td>176.66</td>
<td>2.828</td>
<td>9.03</td>
<td>513.32</td>
<td>19.05</td>
<td>214.16</td>
<td>50.26</td>
</tr>
<tr>
<td>125.00</td>
<td>4.000</td>
<td>12.04</td>
<td>512.07</td>
<td>16.35</td>
<td>159.50</td>
<td>41.73</td>
</tr>
<tr>
<td>88.39</td>
<td>5.657</td>
<td>15.05</td>
<td>510.96</td>
<td>14.08</td>
<td>121.14</td>
<td>36.02</td>
</tr>
<tr>
<td>62.50</td>
<td>8.000</td>
<td>18.06</td>
<td>509.94</td>
<td>12.18</td>
<td>94.29</td>
<td>32.30</td>
</tr>
<tr>
<td>44.19</td>
<td>11.314</td>
<td>21.07</td>
<td>509.16</td>
<td>10.70</td>
<td>75.73</td>
<td>30.04</td>
</tr>
<tr>
<td>31.25</td>
<td>16.000</td>
<td>24.08</td>
<td>508.69</td>
<td>9.66</td>
<td>63.21</td>
<td>28.99</td>
</tr>
<tr>
<td>22.09</td>
<td>22.627</td>
<td>27.09</td>
<td>508.53</td>
<td>9.02</td>
<td>54.93</td>
<td>28.51</td>
</tr>
<tr>
<td>15.63</td>
<td>32.000</td>
<td>30.10</td>
<td>508.51</td>
<td>8.60</td>
<td>49.14</td>
<td>27.42</td>
</tr>
</tbody>
</table>

Table 5.13.2 Means and standard deviations (s.d.) of the measured peak energy factor (PE) and the linear prediction 3 dB bandwidth (B<sub>3dB</sub>) as a function of the actual 3 dB bandwidths of a second-order filter tuned to a 500 Hz resonance frequency. The source waveform is white noise.
### Peak Energy Correlation Matrix: 130 Hz sawtooth and 500 Hz resonant frequency

<table>
<thead>
<tr>
<th></th>
<th>$B_{3\text{dB}}$ (Hz)</th>
<th>$Q_{3\text{dB}}$</th>
<th>$20\log_{10}Q_{3\text{dB}}$</th>
<th>Mean PE (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B_{3\text{dB}}$ (Hz)</td>
<td>1.000</td>
<td>-0.697</td>
<td>-0.909*</td>
<td>-0.947*</td>
</tr>
<tr>
<td>$Q_{3\text{dB}}$</td>
<td>-0.697</td>
<td>1.000</td>
<td>0.909*</td>
<td>0.777*</td>
</tr>
<tr>
<td>$20\log_{10}Q_{3\text{dB}}$</td>
<td>-0.909*</td>
<td>0.909*</td>
<td>1.000</td>
<td>0.964*</td>
</tr>
<tr>
<td>Mean PE (dB)</td>
<td>-0.947*</td>
<td>0.777*</td>
<td>0.964*</td>
<td>1.000</td>
</tr>
</tbody>
</table>

### Linear Prediction Correlation Matrix: 130 Hz sawtooth and 500 Hz resonant frequency

<table>
<thead>
<tr>
<th></th>
<th>$B_{3\text{dB}}$ (Hz)</th>
<th>$Q_{3\text{dB}}$</th>
<th>$20\log_{10}Q_{3\text{dB}}$</th>
<th>Mean $B_{3\text{dB}}$ (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B_{3\text{dB}}$ (Hz)</td>
<td>1.000</td>
<td>-0.679</td>
<td>-0.909*</td>
<td>0.966*</td>
</tr>
<tr>
<td>$Q_{3\text{dB}}$</td>
<td>-0.679</td>
<td>1.000</td>
<td>0.909*</td>
<td>-0.795*</td>
</tr>
<tr>
<td>$20\log_{10}Q_{3\text{dB}}$</td>
<td>-0.909*</td>
<td>0.909*</td>
<td>1.000</td>
<td>-0.975*</td>
</tr>
<tr>
<td>Mean $B_{3\text{dB}}$ (Hz)</td>
<td>0.966*</td>
<td>-0.795*</td>
<td>-0.975*</td>
<td>1.000</td>
</tr>
</tbody>
</table>

### Peak Energy Correlation Matrix: white noise and 500 Hz resonant frequency

<table>
<thead>
<tr>
<th></th>
<th>$B_{3\text{dB}}$ (Hz)</th>
<th>$Q_{3\text{dB}}$</th>
<th>$20\log_{10}Q_{3\text{dB}}$</th>
<th>Mean PE (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B_{3\text{dB}}$ (Hz)</td>
<td>1.000</td>
<td>-0.679</td>
<td>-0.909*</td>
<td>-0.770*</td>
</tr>
<tr>
<td>$Q_{3\text{dB}}$</td>
<td>-0.679</td>
<td>1.000</td>
<td>0.909*</td>
<td>0.957*</td>
</tr>
<tr>
<td>$20\log_{10}Q_{3\text{dB}}$</td>
<td>-0.909*</td>
<td>0.909*</td>
<td>1.000</td>
<td>0.964*</td>
</tr>
<tr>
<td>Mean PE (dB)</td>
<td>-0.770*</td>
<td>0.957*</td>
<td>0.964*</td>
<td>1.000</td>
</tr>
</tbody>
</table>

### Linear Prediction Correlation Matrix: white noise and 500 Hz resonant frequency

<table>
<thead>
<tr>
<th></th>
<th>$B_{3\text{dB}}$ (Hz)</th>
<th>$Q_{3\text{dB}}$</th>
<th>$20\log_{10}Q_{3\text{dB}}$</th>
<th>Mean $B_{3\text{dB}}$ (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B_{3\text{dB}}$ (Hz)</td>
<td>1.000</td>
<td>-0.679</td>
<td>-0.909*</td>
<td>1.000*</td>
</tr>
<tr>
<td>$Q_{3\text{dB}}$</td>
<td>-0.679</td>
<td>1.000</td>
<td>0.909*</td>
<td>-0.673</td>
</tr>
<tr>
<td>$20\log_{10}Q_{3\text{dB}}$</td>
<td>-0.909*</td>
<td>0.909*</td>
<td>1.000</td>
<td>-0.905*</td>
</tr>
<tr>
<td>Mean $B_{3\text{dB}}$ (Hz)</td>
<td>1.000*</td>
<td>-0.673</td>
<td>-0.905*</td>
<td>1.000</td>
</tr>
</tbody>
</table>

**Table 5.14** Correlation matrices of the measured peak energy factor (mean PE) and the 3 dB bandwidths (mean $B_{3\text{dB}}$) obtained through linear prediction. An asterisk indicates a two-tailed significance level of 0.01.
The means and standard deviations of the measured F1s, PE1s, and linear prediction bandwidths are provided in Tables 5.13.1-5.13.2. For the 130 Hz sawtooth signal, the peak energy method clearly locks on to the 4th harmonic (520 Hz) of the periodic waveform once log Q \((20\log_{10}Q_{3\text{dB}})\) exceeds 6 dB. The linear prediction method, on the other hand, seriously underestimates F1 even when damping is taken into account. For example, according to Equation 4.1, the damped frequency \(f_d\) is equal to the resonant frequency \(f_r\) multiplied by a Q-dependent factor: \[f_d = f_r \sqrt{1 - \frac{1}{4Q^2_{3\text{dB}}}}.\] Assuming a 62.5 Hz bandwidth, a not unreasonable value for the B1 of vowels, then \(f_d\) is 499.02 Hz, considerably different from the measured linear prediction F1 of 442.95 Hz given in Table 5.13.1. Typically in linear prediction analysis, speech is differentiated to avoid the excessive low-pass weighting of periodic signals. However, this high-pass filtering would skew upward the already good results of the white noise F1s shown in Table 5.13.2. The peak energy method identifies the white noise F1s fairly well, with the lower F1s being almost certainly the result of damping. A 500 Hz bandwidth, for instance, yields a \(f_d\) of 433.01 Hz when \(f_r\) is equal to 500 Hz. Nevertheless, the standard deviations of the white noise F1s are significantly larger than they are for the linear prediction method.

To evaluate the comparative performance of the peak energy factor (PE) and linear prediction bandwidths, Pearson correlation matrices are presented in Table 5.14. The measured values of the mean PE and the mean LP B3\(_{3\text{dB}}\) all display strong linear relationships with the synthesized values of B3\(_{3\text{dB}}\) and the log Q \((20\log_{10}Q_{3\text{dB}})\). Bandwidth parameters such as the observed LP B3\(_{3\text{dB}}\) and the actual B3\(_{3\text{dB}}\) show high correlations, and exceptionally so for white noise \((\approx 1.0)\). Log amplification parameters such as the observed PE and the actual log Q likewise show very high correlations. Hence the PE
may be viewed as a more auditorily plausible and computationally simpler alternative to linear predicative bandwidth estimation methods. The only drawback to the peak energy method appears to be the large standard deviations of the white noise F1s. On the other hand, the linear prediction method seriously underestimates the F1s of the synthesized periodic waveforms. As a result of the quite different responses to periodic and aperiodic signals, only PEs with like periodicity should be compared with each other. To illustrate, the periodic sawtooth wave with a theoretical 3 dB bandwidth of 125 Hz has a PE of 6.6048 dB, whereas aperiodic white noise with the same theoretical 3 dB bandwidth has a PE of 3.0478 dB.

In this chapter, the analysis methods developed in Chapter 4 were tested in large part by means of synthetically generated signals. One result of particular phonological significance is that aspiration noise is nearly always kept distinct from modulation noise, and vice versa, which is important when noisy voice phonation is present in the speech signal, i.e. harsh voice, harsh whispery voice, and breathy voice. The next chapter applies these analysis methods to actual speech.
NOTES TO CHAPTER 5

1. Two speakers, one man (mdab0) and one woman (felc0) were from the New England dialect region (dr1); the other man (mtas1) and woman (fpas0) were from the Northern dialect region (dr2).

2. The Gaussian approximation is obtained by summing 128 random variables with a rectangular probability distribution. See Schoentgen (2001) for a discussion on the modeling of the probability distribution of jitter.

3. Jitter is set to be a proportion of the standard deviation of the normalized random variable multiplied by 100; the shimmer is fixed as a proportion of the standard deviation of the normalized variable expressed in dB. It is instructive to compare the jitter and shimmer of the synthetic signals so defined with the percent jitter and absolute shimmer calculated according to Equations 3.3 and 3.4. The results are provided in the Table 5.15 below. The jitter values are somewhat similar up to 3%. Beyond this level, the percent jitter increases significantly relative to the jitter as a proportion of the standard deviation. The tendency is even more magnified for shimmer in excess of 1.69 dB. Therefore caution must be exercised when comparing high levels of assigned jitter and shimmer with those obtained from the speech measures. Note that Hillenbrand (1987) set the upper limit of synthetic jitter to 2.6 dB because he observed “the increasingly bizarre perceptual quality of synthesized voice signals as shimmer values approach about 2 dB.” Remark also that the percent shimmer, computed by replacing the periods in Equation 3.3 with the amplitudes, gives a better match than the absolute shimmer. At 128.93 Hz, for example,
the assigned shimmer of 3.57 dB yields a percent shimmer of 57.29% or its decibel equivalent 3.93 dB, which is substantially smaller than the absolute shimmer of 5.73 dB.

<table>
<thead>
<tr>
<th>Jitter (%)</th>
<th>72.88 Hz</th>
<th>128.92 Hz</th>
<th>229.72 Hz</th>
<th>273.18 Hz</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent of s.d.</td>
<td>Percent jitter (%)</td>
<td>0.50</td>
<td>0.57</td>
<td>0.54</td>
</tr>
<tr>
<td>1.00</td>
<td>1.16</td>
<td>1.10</td>
<td>1.15</td>
<td>1.14</td>
</tr>
<tr>
<td>1.50</td>
<td>1.72</td>
<td>1.64</td>
<td>1.69</td>
<td>1.74</td>
</tr>
<tr>
<td>2.00</td>
<td>2.31</td>
<td>2.19</td>
<td>2.25</td>
<td>2.32</td>
</tr>
<tr>
<td>2.50</td>
<td>2.87</td>
<td>2.73</td>
<td>2.82</td>
<td>2.87</td>
</tr>
<tr>
<td>3.00</td>
<td>3.44</td>
<td>3.29</td>
<td>3.39</td>
<td>3.45</td>
</tr>
<tr>
<td>3.50</td>
<td>4.01</td>
<td>3.83</td>
<td>3.95</td>
<td>4.02</td>
</tr>
<tr>
<td>4.00</td>
<td>4.59</td>
<td>4.38</td>
<td>4.52</td>
<td>4.59</td>
</tr>
<tr>
<td>4.50</td>
<td>5.15</td>
<td>4.93</td>
<td>5.09</td>
<td>5.18</td>
</tr>
<tr>
<td>5.00</td>
<td>5.72</td>
<td>5.50</td>
<td>5.65</td>
<td>5.77</td>
</tr>
</tbody>
</table>

| Shimmer (dB) | Log proportion of s.d. | Absolute shimmer (dB) | 0.19 | 0.22 | 0.66 | 0.64 | 0.65 | 0.66 |
|--------------|------------------------|------------------------| 0.94 | 1.14 | 1.11 | 1.13 | 1.14 |
| 1.32 | 1.64 | 1.63 | 1.65 | 1.65 |
| 1.69 | 2.17 | 2.18 | 2.21 | 2.19 |
| 2.07 | 2.78 | 2.84 | 2.87 | 2.81 |
| 2.45 | 3.50 | 3.74 | 3.52 | 3.58 |
| 2.82 | 4.54 | 4.31 | 4.35 | 4.26 |
| 3.20 | 4.59 | 4.99 | 5.36 | 5.16 |
| 3.57 | 5.42 | 5.73 | 5.59 | 6.20 |

**Table 5.15** Jitter (%) and shimmer (dB) as a proportion of the standard deviation of the normalized random variable compared with percent jitter and absolute shimmer calculated according to Equations 3.3 and 3.4.
4. The formant frequencies ($f_r$) from F1 to F4 are those of the adult male speakers in Hillenbrand et al. (1995). The constant formant frequency F5 and all the bandwidths ($B_{3dB}$) are taken from Klatt (1980). See Table 5.16 above.

<table>
<thead>
<tr>
<th>Vowel</th>
<th>$f_r$</th>
<th>$B_{3dB}$</th>
<th>$f_r$</th>
<th>$B_{3dB}$</th>
<th>$f_r$</th>
<th>$B_{3dB}$</th>
<th>$f_r$</th>
<th>$B_{3dB}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>768</td>
<td>130</td>
<td>588</td>
<td>70</td>
<td>378</td>
<td>65</td>
<td>342</td>
<td>60</td>
</tr>
<tr>
<td>F2</td>
<td>1333</td>
<td>70</td>
<td>1952</td>
<td>150</td>
<td>997</td>
<td>110</td>
<td>2322</td>
<td>200</td>
</tr>
<tr>
<td>F3</td>
<td>2522</td>
<td>160</td>
<td>2601</td>
<td>320</td>
<td>2343</td>
<td>140</td>
<td>3000</td>
<td>400</td>
</tr>
<tr>
<td>F4</td>
<td>3687</td>
<td>250</td>
<td>3624</td>
<td>250</td>
<td>3357</td>
<td>250</td>
<td>3657</td>
<td>250</td>
</tr>
<tr>
<td>F5</td>
<td>3750</td>
<td>200</td>
<td>3750</td>
<td>200</td>
<td>3750</td>
<td>200</td>
<td>3750</td>
<td>200</td>
</tr>
</tbody>
</table>

**Table 5.16** Formant frequencies and bandwidths used in the synthesis of men’s vowels.
CHAPTER 6

PHONETIC ANALYSIS OF THE PHONATION TYPES

6.1 Glottal noise: glottal stop, whisper, breath

6.1.1 Standard Thai

As was pointed out in Chapter 2 (Section 2.4.2.1), many languages contrast the

glottal stop /ʔ/ and voicelessness /h/. Standard Thai appears to be such a language. In the

Handbook of the International Phonetic Association, Tingsabadh and Abramson (1999:

148) list the near-minimal pair ?ān ‘saddle’ and hān ‘to divide’. The acoustic waveforms

and spectrograms of the two tokens are presented in Figure 6.1. The F1s and PE1s of the

initial, mid, and final sampling points are provided in Table 6.1. To achieve more

certainty with regard to the frequency analysis, the location of F1 is cross-checked

against the spectrographic display of the waveform editor. If a mismatch occurs, the

search range for F1 is adjusted accordingly. In addition, second formant frequencies (F2)

and peak energy factors (PE2) are given for comparison purposes. With the exception of

the final measure near the onset of the vowel, the PE1 of /h/ is much lower than for /ʔ/, as

expected in view of the greater damping caused by a more open glottis. Although PE2

should be affected by the glottal damping as well, albeit to a lesser degree (see discussion

in Section 2.4.1), the midpoint PE2 of /h/ is much greater than that of /ʔ/, showing the

relative independence of PE1 and PE2. The midpoint formant measures are evidently

better indicators of the anticipated results than the initial and final ones, most likely

because of its greater robustness against coarticulatory effects.
Figure 6.1 Standard Thai ฎำน ‘saddle’ (above), หำน ‘to divide’ (below).
The Northeast Caucasian language Avar exhibits an opposition between postvocalic /ʔ/ and /h/. There is also a phonological contrast between voiceless and voiced pharyngeals /h, ŋ/ (Charachidzé, 1981: 28; see also Hewitt, 2004: 50, 270-272 and references therein). Postvocalic sound distinctions of a woman speaker are illustrated in the 1991 HyperCard version of the UCLA Sounds of the World’s Languages: wabaʔ ‘cholera’, mah ‘bundle’, mah ‘odor’, maŋ ‘nail’. The voiceless and voiced “epiglottal” fricatives h and ŋ most likely represent phonetic transcriptions of phonological /h/ and /ŋ/, respectively. The spectrograms of the four words are shown in Figures 6.2.1-6.2.2. To ensure that the time window takes into account the entire signal, 100 ms of silence are added to the beginning and end of this UCLA speech sample as well as the two others used in this study (Quechua, Mpi).

### Table 6.1 Duration and formant parameters for prevocalic /ʔ, h/ in Standard Thai.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Consonant</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration</td>
<td>?</td>
<td>139.7-147.1: 7.4 ms</td>
<td>100.5-257.2: 156.7 ms</td>
<td></td>
</tr>
<tr>
<td></td>
<td>h</td>
<td>486.77</td>
<td>718.87</td>
<td>905.72</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1031.42</td>
<td>1016.63</td>
<td>688.39</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>?</td>
<td>0.93383</td>
<td>0.88283</td>
<td>1.57064</td>
</tr>
<tr>
<td></td>
<td>h</td>
<td>0.00517</td>
<td>0.27898</td>
<td>2.08410</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>?</td>
<td>1.191.65</td>
<td>1191.65</td>
<td>1226.57</td>
</tr>
<tr>
<td></td>
<td>h</td>
<td>1545.38</td>
<td>1523.23</td>
<td>1458.65</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>?</td>
<td>1.71421</td>
<td>1.38450</td>
<td>1.62046</td>
</tr>
<tr>
<td></td>
<td>h</td>
<td>0.05875</td>
<td>2.84245</td>
<td>3.82541</td>
</tr>
</tbody>
</table>

6.1.2 Avar

The Northeast Caucasian language Avar exhibits an opposition between postvocalic /ʔ/ and /h/. There is also a phonological contrast between voiceless and voiced pharyngeals /h, ŋ/ (Charachidzé, 1981: 28; see also Hewitt, 2004: 50, 270-272 and references therein). Postvocalic sound distinctions of a woman speaker are illustrated in the 1991 HyperCard version of the UCLA Sounds of the World’s Languages: wabaʔ ‘cholera’, mah ‘bundle’, mah ‘odor’, maŋ ‘nail’. The voiceless and voiced “epiglottal” fricatives h and ŋ most likely represent phonetic transcriptions of phonological /h/ and /ŋ/, respectively. The spectrograms of the four words are shown in Figures 6.2.1-6.2.2. To ensure that the time window takes into account the entire signal, 100 ms of silence are added to the beginning and end of this UCLA speech sample as well as the two others used in this study (Quechua, Mpi).
Figure 6.2.1 Avar waba? ‘cholera’ (above), mah ‘bundle’ (below).
Figure 6.2.2 Avar ман ‘odor’ (above), маf ‘nail’ (below).
As seen in the audio signal and spectrogram of waba? (Figure 6.2.1), the release of the glottal stop ? occurs at about 900 ms and is followed by a schwa-like vowel. In Figure 6.2.2, the epiglottal fricative ? of ma? begins to devoice at 600 ms, then a distinct popping sound is heard at 700 ms, the latter apparently consisting of the release of the strong or pharyngealized glottal stop ? (see the discussion in Section 2.4.2.3 of Chapter 2). The results of the analysis are given in Table 6.2 above.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Consonant</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration</td>
<td>?</td>
<td>864.8-882.1: 17.3 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>?</td>
<td>667.1-716.5: 49.4 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>h</td>
<td>588.0-877.2: 289.2 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>h</td>
<td>481.8-827.7: 345.9 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>?</td>
<td>750.70</td>
<td>729.32</td>
<td>729.32</td>
</tr>
<tr>
<td></td>
<td>?</td>
<td>750.70</td>
<td>718.87</td>
<td>718.87</td>
</tr>
<tr>
<td></td>
<td>h</td>
<td>892.73</td>
<td>973.53</td>
<td>879.93</td>
</tr>
<tr>
<td></td>
<td>h</td>
<td>783.93</td>
<td>867.32</td>
<td>932.26</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>?</td>
<td>3.70135</td>
<td>3.99639</td>
<td>4.35208</td>
</tr>
<tr>
<td></td>
<td>?</td>
<td>2.43433</td>
<td>3.40718</td>
<td>4.12072</td>
</tr>
<tr>
<td></td>
<td>h</td>
<td>3.36086</td>
<td>1.86915</td>
<td>1.51445</td>
</tr>
<tr>
<td></td>
<td>h</td>
<td>1.96685</td>
<td>2.56840</td>
<td>5.48033</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>?</td>
<td>1545.38</td>
<td>1545.38</td>
<td>1545.38</td>
</tr>
<tr>
<td></td>
<td>?</td>
<td>1226.57</td>
<td>1174.57</td>
<td>1174.57</td>
</tr>
<tr>
<td></td>
<td>h</td>
<td>1523.23</td>
<td>1590.67</td>
<td>1567.86</td>
</tr>
<tr>
<td></td>
<td>h</td>
<td>1376.78</td>
<td>2033.27</td>
<td>1811.43</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>?</td>
<td>3.72305</td>
<td>3.91224</td>
<td>3.72915</td>
</tr>
<tr>
<td></td>
<td>?</td>
<td>2.67405</td>
<td>3.50658</td>
<td>3.55952</td>
</tr>
<tr>
<td></td>
<td>h</td>
<td>3.80068</td>
<td>3.31412</td>
<td>2.84813</td>
</tr>
<tr>
<td></td>
<td>h</td>
<td>1.73111</td>
<td>4.12577</td>
<td>4.60676</td>
</tr>
</tbody>
</table>

Table 6.2 Duration and formant parameters for postvocalic /?, h, ?, h/ in Avar.
PE1 measures at midpoint clearly oppose the two varieties of glottal stops (ʔ: 3.99639 dB; h: 3.40718 dB) to the voiceless fricatives (h: 1.86915 dB; h: 2.56840 dB), the larger values of the glottal stops signaling a narrower glottal orifice. In addition, the smaller PE1 value of h relative to h may be a cue for the contrast breath (h) vs. whisper (h), but more evidence is needed to substantiate this possibility. Note also that the F2s of ʔ (1545.38 Hz) and h (1590.67 Hz) probably indicate a central position of the hindtongue (tongue body and root) during the production of these glottal noise sounds, as compared to a front position for h (2033.27 Hz) and a back position for ʔ (1174.57 Hz).

6.1.3 Quechua

Quechua contrasts voiceless, aspirated voiceless, and ejective obstruents. Nine words produced by a male Quechua speaker are taken from the current UCLA Phonetics Lab Database and illustrated in Table 6.3. Because the glottis is completely sealed as a result of glottalic initiation, the PE1 should be high during the supraglottal release phase of the

<table>
<thead>
<tr>
<th>Quechua</th>
<th>Palato-alveolar</th>
<th>Velar</th>
<th>Uvular</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voiceless</td>
<td>tʃ’aka</td>
<td>kujui</td>
<td>qaʔu</td>
</tr>
<tr>
<td>Aspirated voiceless</td>
<td>tʃʰaka</td>
<td>kʰujui</td>
<td>qʰaʔu</td>
</tr>
<tr>
<td>Ejective</td>
<td>tʃ’aka</td>
<td>k’ujui</td>
<td>q’aʔu</td>
</tr>
</tbody>
</table>

Table 6.3 Quechua words in the current UCLA Phonetics Lab Database illustrating the contrasts among voiceless, aspirated voiceless, and ejective stops at three places of articulation.
ejective in comparison to its value during the release burst of a pulmonic stop made with an open glottis (voiceless, aspirated voiceless). To test this hypothesis, the three-way contrasts at each place of articulation are examined.

Figure 6.3.1 Quechua qaũ ‘tongue’.
Figure 6.3.2 Quechua qʰaʎu ‘shawl’ (above), q’aʎu ‘tomato sauce’ (below).
Data for the uvular stops are presented in Table 6.4.1. For the aspirated voiceless stop /qʰ/, measures are made only over the subsegment comprising the release burst 134.8-176.4: 41.6 ms; the aspiration subsegment of qʰaʃu extending from 178.9 to 289.1 ms in the spectrogram and audio signal of Figure 6.3.2 is not taken into account. Likewise, only the supraglottal release burst of /q’/ is measured (137.2-203.4: 66.2 ms).

Comparison of the values in Table 6.4.1 shows that the PE1 of the ejective burst is considerably larger than those of the voiceless and aspirated voiceless bursts at the initial, mid, and final frames. On the other hand, there does not appear to be a significant difference between the releases of the voiceless and the aspirated voiceless stops.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Consonant</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration</td>
<td>q’</td>
<td>137.2-203.4: 66.2 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>q</td>
<td>110.3-132.3: 22 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>qʰ</td>
<td>134.8-176.4: 41.6 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>q’</td>
<td>750.70</td>
<td>538.54</td>
<td>698.40</td>
</tr>
<tr>
<td></td>
<td>q</td>
<td>640.44</td>
<td>659.20</td>
<td>446.37</td>
</tr>
<tr>
<td></td>
<td>qʰ</td>
<td>546.38</td>
<td>403.45</td>
<td>562.39</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>q’</td>
<td>2.58155</td>
<td>3.59493</td>
<td>8.67580</td>
</tr>
<tr>
<td></td>
<td>q</td>
<td>0.80763</td>
<td>1.55708</td>
<td>2.40556</td>
</tr>
<tr>
<td></td>
<td>qʰ</td>
<td>0.82959</td>
<td>2.01208</td>
<td>1.42709</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>q’</td>
<td>1174.57</td>
<td>1837.78</td>
<td>1637.28</td>
</tr>
<tr>
<td></td>
<td>q</td>
<td>1376.78</td>
<td>1357.04</td>
<td>1357.04</td>
</tr>
<tr>
<td></td>
<td>qʰ</td>
<td>1262.51</td>
<td>1226.57</td>
<td>1226.57</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>q’</td>
<td>2.95177</td>
<td>2.81710</td>
<td>4.76878</td>
</tr>
<tr>
<td></td>
<td>q</td>
<td>2.88719</td>
<td>2.14996</td>
<td>2.02087</td>
</tr>
<tr>
<td></td>
<td>qʰ</td>
<td>3.72106</td>
<td>3.86786</td>
<td>3.39476</td>
</tr>
</tbody>
</table>

Table 6.4.1 Duration and formant parameters for the release bursts of /q, q’, qʰ/ in Quechua.
Figure 6.3.3 Quechua kujui ‘to move’ (above), ḱujui ‘to whistle’ (below).
Figure 6.3.4 Quechua k’ujui ‘to twist’.
Data for the velar stops are given in Table 6.4.2. Similarly to the uvular stops, the PE1s of ejectives are greater than those of the voiceless and aspirated voiceless stops, with one exception: the initial measures show a slightly higher PE1 value for the unaspirated velar. Remark that the bursts of the pulmonic aspirated velars have lower PE1s than the corresponding pulmonic unaspirated velars, an indication that the glottal opening of aspirated stops is wider than those of aspirated stops during the release phase (cf. Löfqvist and Yoshioka, 1981). Nonetheless, this distinction between breath and whisper is far from systematic because it does not hold at all for the uvular series discussed above.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Consonant</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration</td>
<td>k'</td>
<td>137.2-225.4: 88.2 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>k</td>
<td>147.0-191.1: 44.1 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>kʰ</td>
<td>147.0-186.2: 39.2 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>k'</td>
<td>397.67</td>
<td>380.81</td>
<td>375.35</td>
</tr>
<tr>
<td></td>
<td>k</td>
<td>472.91</td>
<td>493.85</td>
<td>311.10</td>
</tr>
<tr>
<td></td>
<td>kʰ</td>
<td>421.31</td>
<td>302.25</td>
<td>293.64</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>k'</td>
<td>1.33826</td>
<td>2.76190</td>
<td>4.67424</td>
</tr>
<tr>
<td></td>
<td>k</td>
<td>1.50379</td>
<td>2.04215</td>
<td>2.48669</td>
</tr>
<tr>
<td></td>
<td>kʰ</td>
<td>0.69868</td>
<td>1.53787</td>
<td>1.71838</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>k'</td>
<td>854.88</td>
<td>892.73</td>
<td>1002.06</td>
</tr>
<tr>
<td></td>
<td>k</td>
<td>688.39</td>
<td>668.79</td>
<td>659.20</td>
</tr>
<tr>
<td></td>
<td>kʰ</td>
<td>918.89</td>
<td>892.73</td>
<td>892.73</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>k'</td>
<td>2.95432</td>
<td>2.41133</td>
<td>7.27446</td>
</tr>
<tr>
<td></td>
<td>k</td>
<td>3.16105</td>
<td>2.67677</td>
<td>2.77861</td>
</tr>
<tr>
<td></td>
<td>kʰ</td>
<td>1.99186</td>
<td>2.06911</td>
<td>2.00068</td>
</tr>
</tbody>
</table>

Table 6.4.2 Duration and formant parameters for the release bursts of /k, k’, kʰ/ in Quechua.
Figure 6.3.5 Quechua tʃaka ‘bridge’ (above), tʃʰaka ‘large ant’ (below).
Figure 6.3.6 Quechua tj’aka ‘hoarse’.
Formant parameters for the first or stop subsegment of the palato-alveolar affricates are provided in Table 6.4.3. Apart from the fact that the PE1 of the aspirated affricate is somewhat higher than its values for the other series, there appears to be no consistent pattern of differences between them. Hence it is possible that the initiation of the stop portion of the affricate is pulmonic for all three series. To check whether this is effectively the case, physiological measures would need to be assessed.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Consonant</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration</td>
<td>t(ʃ)'</td>
<td>144.6-181.3: 36.7 ms</td>
<td>115.2-144.6: 29.4 ms</td>
<td>127.4-139.7: 12.3 ms</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>t(ʃ)'</td>
<td>708.56</td>
<td>729.32</td>
<td>446.37</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>t(ʃ)'</td>
<td>1.55513</td>
<td>1.39516</td>
<td>1.65511</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>t(ʃ)'</td>
<td>1280.88</td>
<td>1244.41</td>
<td>1244.41</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>t(ʃ)'</td>
<td>2.10448</td>
<td>1.55932</td>
<td>1.49403</td>
</tr>
</tbody>
</table>

Table 6.4.3 Duration and formant parameters for the release bursts of the first subsegment of the palato-alveolar affricate /t(ʃ), t(ʃ)', t(ʃ)'h/ in Quechua.
Formant parameters for the second or fricative subsegment of the palato-alveolar affricates are presented in Table 6.4.4. Excepting final /tʰ/, the PE1s of the fricative portion of the affricate are highest for the ejective (glottal stop), lowest for the aspirated (breath), and intermediate for the unaspirated (whisper) sounds. Consequently, the results nearly parallel those of the velar series.

6.1.4 Summary of the results for glottal noise

The first formant peak energy factor PE1 appears to perform fairly well as a cue for glottal aperture, as would be expected from the discussion of Flanagan (1972) and Stevens (1998) in Section 2.4.1. First formant damping caused by shunt losses is small

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Consonant</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration</td>
<td>(t)ʃʰ</td>
<td>183.8-257.2: 73.4 ms</td>
<td>147.0-210.7: 63.7 ms</td>
<td>142.1-249.9: 107.8 ms</td>
</tr>
<tr>
<td></td>
<td>(t)ʃ</td>
<td>622.21</td>
<td>613.29</td>
<td>604.49</td>
</tr>
<tr>
<td></td>
<td>(t)ʃʰ</td>
<td>493.85</td>
<td>369.97</td>
<td>508.32</td>
</tr>
<tr>
<td></td>
<td>(t)ʃʰ</td>
<td>729.32</td>
<td>546.38</td>
<td>613.29</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>(t)ʃʰ</td>
<td>2.83098</td>
<td>4.68520</td>
<td>4.46430</td>
</tr>
<tr>
<td></td>
<td>(t)ʃ</td>
<td>2.08810</td>
<td>1.57061</td>
<td>1.59171</td>
</tr>
<tr>
<td></td>
<td>(t)ʃʰ</td>
<td>1.29690</td>
<td>0.69040</td>
<td>5.08871</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>(t)ʃʰ</td>
<td>1208.99</td>
<td>1046.42</td>
<td>1046.42</td>
</tr>
<tr>
<td></td>
<td>(t)ʃ</td>
<td>1280.88</td>
<td>1919.15</td>
<td>1785.46</td>
</tr>
<tr>
<td></td>
<td>(t)ʃʰ</td>
<td>1244.41</td>
<td>1709.77</td>
<td>1046.42</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>(t)ʃʰ</td>
<td>2.41872</td>
<td>2.77731</td>
<td>4.98964</td>
</tr>
<tr>
<td></td>
<td>(t)ʃ</td>
<td>1.42927</td>
<td>2.19404</td>
<td>1.09918</td>
</tr>
<tr>
<td></td>
<td>(t)ʃʰ</td>
<td>1.67759</td>
<td>0.63566</td>
<td>4.01777</td>
</tr>
</tbody>
</table>

Table 6.4.4 Duration and formant parameters for the second subsegment of the palato-alveolar affricate /(tʃ, (tʃʰ, (tʃʰ/) in Quechua.
when the glottis is narrowly constricted or closed during the production of a *glottal stop*. Larger amounts of damping result from the narrowly open glottis of *whisper*, and additionally so from the widely open glottis of *breath*. An increase in glottal damping leads to a broadening mainly of the first formant bandwidth and a corresponding reduction in PE1.

Phonemic glottal stop */ʔ/* nearly always displays a higher PE1 value than does phonemic voicelessness */h/* (*breath* or *whisper*), whether in prevocalic or postvocalic positions (Standard Thai and Avar, respectively). This trend is particularly strong at the midpoint where there is greater robustness against coarticulatory effects. Quechua ejectives, generated by an initiatory glottal stop, likewise show higher PE1s during the supraglottal release phase than do unaspirated and aspirated pulmonic stops produced with an open glottis. There is some tendency for the unaspirated sounds to exhibit higher PE1s during release than the aspirated ones, indicating a possible opposition between *whisper* and *breath* at the glottal source.

One aspect of using the PE1 as a cue for glottal aperture has only been touched upon. It was mentioned in Chapter 2 that the area of the nasal port must remain constant in order for a valid comparison between the different types of glottal noise to be made. The reason is that increasing nasalization also broadens the F1 bandwidth (Delattre, 1954; House and Stevens, 1956; Fant, 1960; Fujimura, 1960; Carré, 1975; Chen 1997). Moreover, Matisoff (1975) points out that the phonetic similarity between nasalization and aspiration, deemed “rhinoglottophilia”, can lead to phonological processes that seem inexplicable at first glance (see also Ohala, 1975 and Blevins, 2004: 135-136 for recent
6.2 Pure voice: tense voice and plain voice in Mpi

As defined in Chapter 2, pure voice is characterized by a periodic glottal waveform with negligible aspiration or modulation noise. Tense voice has a spectral tilt significantly flatter than plain voice, plain voice has a spectral tilt lying between lax voice and tense voice, whereas lax voice has a spectral tilt significantly steeper than plain voice. The contrast between plain voice and tense voice is illustrated in the current UCLA Phonetics Lab CD by the Tibeto-Burman language Mpi (see Blankenship, 1997 for additional information on the language and recording conditions). The twelve words in Table 6.5 combine the two phonation types of /i/ with six tones. The glottal source and formant measures of the twelve words are given in Appendix A (Tables A.1-A.12). If there is more than one subsegment, an asterisk (*) indicates the one with the maximum energy.

<table>
<thead>
<tr>
<th>Tone</th>
<th>Mpi word</th>
<th>Plain voice</th>
<th>Mpi word</th>
<th>Tense voice</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Level</td>
<td>‘four’</td>
<td>si</td>
<td>(name)</td>
<td>si</td>
</tr>
<tr>
<td>High Falling</td>
<td>‘to die’</td>
<td>si</td>
<td>(name)</td>
<td>si</td>
</tr>
<tr>
<td>Mid Level</td>
<td>(a color)</td>
<td>si</td>
<td>(classifier)</td>
<td>si</td>
</tr>
<tr>
<td>Mid Rising</td>
<td>‘to roll rope’</td>
<td>si</td>
<td>‘to smoke’</td>
<td>si</td>
</tr>
<tr>
<td>Low Level</td>
<td>‘blood’</td>
<td>si</td>
<td>‘seven’</td>
<td>si</td>
</tr>
<tr>
<td>Low Rising</td>
<td>‘to be putrid’</td>
<td>si</td>
<td>to be dried up</td>
<td>si</td>
</tr>
</tbody>
</table>

Table 6.5 Mpi words in the current UCLA Phonetics Lab Database illustrating the contrast between plain and tense voice. The twelve words are produced by a male speaker.
To provide an overview of the differences between tense and plain voice, the parameters are averaged across the six tones as well as the initial and mid measures of the first subsegment if there is more than one (Table 6.6). The final measures of vowels often diverge quite considerably from the initial and midpoint ones. Typically the log pACC and the HNR are lowered while the spectral tilt becomes fairly steep (see the discussion at the end of Section 6.3.2). Hence the final measures are not included. A one-way analysis of variance is also calculated (plain voice N = 12; tense voice N = 12).

<table>
<thead>
<tr>
<th></th>
<th>Mpi: man</th>
<th>/i/</th>
<th>Mean</th>
<th>S.D.</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>F0 (Hz)</td>
<td></td>
<td></td>
<td>plain</td>
<td>138.00</td>
<td>14.86</td>
<td>0.172</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>tense</td>
<td>140.70</td>
<td>16.97</td>
<td></td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td></td>
<td></td>
<td>plain</td>
<td>-0.9555</td>
<td>0.1723</td>
<td>0.162</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>tense</td>
<td>-0.9880</td>
<td>0.2204</td>
<td></td>
</tr>
<tr>
<td>HNR (dB)</td>
<td></td>
<td></td>
<td>plain</td>
<td>23.330</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>tense</td>
<td>22.452</td>
<td>3.041</td>
<td></td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td></td>
<td></td>
<td>plain</td>
<td>-4.421</td>
<td>0.259</td>
<td>23.908</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>tense</td>
<td>-3.904</td>
<td>0.260</td>
<td></td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td></td>
<td></td>
<td>plain</td>
<td>-14.303</td>
<td>1.710</td>
<td>0.169</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>tense</td>
<td>-14.685</td>
<td>2.718</td>
<td></td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td></td>
<td></td>
<td>plain</td>
<td>375.36</td>
<td>70.49</td>
<td>0.951</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>tense</td>
<td>405.95</td>
<td>82.70</td>
<td></td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td></td>
<td></td>
<td>plain</td>
<td>6.41004</td>
<td>0.90547</td>
<td>1.412</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>tense</td>
<td>6.75117</td>
<td>0.41130</td>
<td></td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td></td>
<td></td>
<td>plain</td>
<td>2010.52</td>
<td>83.06</td>
<td>25.097</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>tense</td>
<td>2223.04</td>
<td>121.23</td>
<td></td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td></td>
<td></td>
<td>plain</td>
<td>5.40223</td>
<td>1.18117</td>
<td>1.433</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>tense</td>
<td>4.77807</td>
<td>1.36664</td>
<td></td>
</tr>
</tbody>
</table>

**Table 6.6** Means and standard deviations of the glottal source and formant parameters of Mpi male /i/, averaged across the six tones as well as the initial and mid measures of the first subsegment. The results of the one-way ANOVA are given in the last two columns.
Because neither the log pACCs nor the HNRs are diminished, Mpi plain and tense voice are clearly instances of pure voice. For example, when a 128.93 Hz glottal waveform is synthesized with normal levels of modulation noise, the log pACC is –1.04 dB (see Table 5.5 in Chapter 5). This value is very similar to the averages of plain voice (–0.9555 dB, 138.00 Hz) and tense voice (–0.9880 dB, 140.70 Hz). Likewise, the mean HNRs of plain and tense voice are appreciably higher than the posited upper threshold range of 15-18 dB for aspiration noise (see the end of Section 5.4.2).

Within the category of pure voice it appears that the relevant contrast is between plain voice and tense voice. Plain voice is characterized by a mean harmonic slope of –4.421 dB/octave, tense voice –3.904 dB/octave. In Table 5.12 the harmonic slope of /i/ is –4.19 dB/octave for the natural adult male vowel, –5.22 dB/octave for the synthetic adult male vowel, and –4.82 dB/octave for the adult female vowel. On the whole, the plain voice slope of –4.421 dB/octave is closer to these values than the flatter tense voice harmonic slope of –3.904 dB/octave. As a consequence, the statistically significant difference between the two harmonic slopes (Sig. < 0.05) very probably cues the phonological opposition plain voice vs. tense voice.

Compared to the wide-band harmonic slope, the narrow-band spectral tilt H1–H2 performs very poorly in keeping plain and tense voice distinct, as shown by the low nonsignificant F ratio (0.169) in Table 6.6. In addition, if the spectral tilts of plain and tense voice are compared pairwise at each of the three sampling points of the first subsegment (initial, mid, final), the harmonic slope gives 18 correct results out of the 18 possible cases (100%), whereas H1–H2 yields only 12 (67%). Because the phonemic distinction between Mpi plain and tense vowels appears to rely heavily on spectral tilt, it
is likely that the H1–H2 parameter plays a fairly minor role in auditory speech processing. Nevertheless, the performance of H1–H2 improves somewhat in Gujarati and Jalapa Mazatec, as will be seen in remaining sections. Furthermore, the contrast in spectral tilt is not the only significant difference between Mpi plain and tense vowels: the tense /i/ has a higher F2 than the plain /i/. The most important theoretical consideration against the narrow-band spectral tilt H1–H2 is that the glottal bandwidth (GBW) corresponds to the effective bandwidth of the *entire* speech spectrum (see discussion in Section 2.4.2.4 of Chapter 2). This suggests that the wide-band harmonic slope, estimated simultaneously with the HNR by the ASHE method, is a more appropriate measure of spectral tilt than the narrow-band spectral tilt H1–H2.

6.3 Noisy voice: harsh voice, harsh whispery voice, breathy voice

6.3.1 Gujarati

The noisy voice phonation types are characterized by a periodic glottal source mixed with a large amount of aperiodicity. The aperiodicity consists either of modulation noise (*harsh voice*) or aspiration noise (*breathy voice*). A low harmonics-to-noise ratio (HNR < 15-18 dB) corresponds to significant aspiration noise. Gujarati exhibits a marginal phonemic contrast between pure and breathy voice. The breathy vowels are the result of contracting a word with an original /h/, realized [h] or [fi] and written в in Gujarati script. To illustrate, the word вүл ‘outside’ has three alternative pronunciations [bөхөг], [бар], or [баг] (Dave, 1967). If the last realization is employed, then the word forms a minimal pair with вүл ‘twelve’ [бар]. Our principal informant was a twenty-nine year old male graduate student in computer science, a native speaker of Ahmedabad
Gujarati with some linguistic training. From a list of candidate words drawn from Pandit (1957), Dave (1967), and Fischer-Jørgensen (1967), he selected the ones most likely to be stable minimal pairs. They are presented in Table 6.7.

<table>
<thead>
<tr>
<th>Gujarati word</th>
<th>Pure voice</th>
<th>Gujarati word</th>
<th>Breathy voice</th>
</tr>
</thead>
<tbody>
<tr>
<td>પાઠ ‘obligation’</td>
<td>pad</td>
<td>પાઠ ‘mountain’</td>
<td>pad</td>
</tr>
<tr>
<td>પાર ‘twelve’</td>
<td>bar</td>
<td>પાર ‘outside’</td>
<td>bar</td>
</tr>
<tr>
<td>માન ‘goods, stuff, money’</td>
<td>mal</td>
<td>માન ‘rejoice, enjoy’</td>
<td>mal</td>
</tr>
<tr>
<td>ોખ ‘that one’ (masc.)</td>
<td>pelo</td>
<td>ોખ ‘first’ (masc.)</td>
<td>pelo</td>
</tr>
<tr>
<td>મેલ ‘dirt’</td>
<td>mel</td>
<td>મેલ ‘palace’</td>
<td>mel</td>
</tr>
<tr>
<td>મેર ‘revenge, animosity’</td>
<td>wer</td>
<td>મેર ‘sawdust’</td>
<td>wer</td>
</tr>
</tbody>
</table>

Table 6.7 Gujarati minimal pairs illustrating the contrast between pure and breathy voice.

The words were recorded by two native speakers of Gujarati, one male and one female, each of whom was an eighteen year old Indiana University freshman recently arrived from Gujarat. The twelve lexical items were repeated three times, both in isolation and in a carrier sentence. The first word in the isolated sequence was chosen for acoustic analysis. DAT recordings were made in an anechoic chamber with a flat-response dynamic microphone. The digital audio on the tape was then transferred to 44.1 kHz PCM files. The female glottal source and formant measures of the twelve words are shown in Appendix B (Tables B.1.1-B.1.12); the male measures are also given in Appendix B (Tables B.2.1-B.2.12). If there is more than one subsegment, an asterisk (*) indicates the one with the maximum energy.
To illustrate the differences between pure and breathy voice for each speaker, the initial and mid measures of the first subsegment are averaged across the three tokens of each vowel. A one-way analysis of variance is computed (pure voice $N = 6$; breathy voice $N = 6$). The woman’s results are provided in Tables 6.8.1-6.8.2, the man’s in Tables 6.9.1-6.9.2.

<table>
<thead>
<tr>
<th>Gujarati: woman</th>
<th>/a/</th>
<th>Mean</th>
<th>S.D.</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>F0 (Hz)</td>
<td></td>
<td>pure</td>
<td>268.87</td>
<td>12.54</td>
<td>4.673</td>
</tr>
<tr>
<td></td>
<td></td>
<td>breathy</td>
<td>253.77</td>
<td>11.63</td>
<td></td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td></td>
<td>pure</td>
<td>–0.6912</td>
<td>0.3881</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td></td>
<td>breathy</td>
<td>–0.6499</td>
<td>0.2000</td>
<td></td>
</tr>
<tr>
<td>HNR (dB)</td>
<td></td>
<td>pure</td>
<td>23.330</td>
<td>0.000</td>
<td>N.A.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>breathy</td>
<td>23.330</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td></td>
<td>pure</td>
<td>–3.253</td>
<td>0.472</td>
<td>5.484</td>
</tr>
<tr>
<td></td>
<td></td>
<td>breathy</td>
<td>–3.951</td>
<td>0.557</td>
<td></td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td></td>
<td>pure</td>
<td>4.585</td>
<td>2.523</td>
<td>1.809</td>
</tr>
<tr>
<td></td>
<td></td>
<td>breathy</td>
<td>6.134</td>
<td>1.261</td>
<td></td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td></td>
<td>pure</td>
<td>1060.63</td>
<td>63.00</td>
<td>2.336</td>
</tr>
<tr>
<td></td>
<td></td>
<td>breathy</td>
<td>1017.02</td>
<td>30.27</td>
<td></td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td></td>
<td>pure</td>
<td>4.96563</td>
<td>1.99332</td>
<td>0.684</td>
</tr>
<tr>
<td></td>
<td></td>
<td>breathy</td>
<td>6.03312</td>
<td>2.45495</td>
<td></td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td></td>
<td>pure</td>
<td>1383.44</td>
<td>192.15</td>
<td>0.036</td>
</tr>
<tr>
<td></td>
<td></td>
<td>breathy</td>
<td>1366.68</td>
<td>102.81</td>
<td></td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td></td>
<td>pure</td>
<td>6.09827</td>
<td>2.62111</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td></td>
<td>breathy</td>
<td>5.85486</td>
<td>2.53641</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.8.1 Means and standard deviations of the glottal source and formant parameters for Gujarati female /a/, averaged across the three tokens of the vowel as well as the initial and mid measures of the first subsegment. The results of the one-way ANOVA are given in the last two columns (N.A. means not applicable).
There is a clear difference in behavior between the woman’s vowels /a/ and /e/.

The phonologically breathy vowel /ä/ is kept separate from the pure vowel /a/ only through a significant distinction in harmonic slope (Sig. < 0.05), the HNRs all being identically elevated (23.330 dB). As a result, the phonemic opposition depends on the contrast between two pure voice phonation types: lax voice (−3.951 dB/octave) vs. plain voice (−3.253 dB/octave). The latter value is very close to the natural adult female slope of −3.27 dB/octave for [a] shown in Table 5.12. Note that the natural adult male slope is −3.49 dB/octave and the synthetic adult male slope is −3.53 dB/octave for the same

<table>
<thead>
<tr>
<th>Gujarati: woman /e/</th>
<th>Mean</th>
<th>S.D.</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>F0 (Hz)</td>
<td>pure breathy</td>
<td>266.19</td>
<td>9.32</td>
<td>0.282</td>
</tr>
<tr>
<td></td>
<td>262.54</td>
<td></td>
<td>14.01</td>
<td></td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>pure breathy</td>
<td>−0.4851</td>
<td>0.1010</td>
<td>0.227</td>
</tr>
<tr>
<td></td>
<td>−0.5190</td>
<td></td>
<td>0.1421</td>
<td></td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>pure breathy</td>
<td>19.316</td>
<td>6.291</td>
<td>10.910</td>
</tr>
<tr>
<td></td>
<td>9.533</td>
<td></td>
<td>3.615</td>
<td></td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>pure breathy</td>
<td>−5.205</td>
<td>0.450</td>
<td>14.841</td>
</tr>
<tr>
<td></td>
<td>−6.765</td>
<td></td>
<td>0.884</td>
<td></td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>pure breathy</td>
<td>−1.068</td>
<td>3.491</td>
<td>16.264</td>
</tr>
<tr>
<td></td>
<td>5.514</td>
<td></td>
<td>1.948</td>
<td></td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>pure breathy</td>
<td>958.79</td>
<td>334.61</td>
<td>0.216</td>
</tr>
<tr>
<td></td>
<td>875.62</td>
<td></td>
<td>283.15</td>
<td></td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>pure breathy</td>
<td>5.62854</td>
<td>1.57211</td>
<td>0.785</td>
</tr>
<tr>
<td></td>
<td>6.25517</td>
<td></td>
<td>0.72698</td>
<td></td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>pure breathy</td>
<td>2511.28</td>
<td>162.98</td>
<td>3.241</td>
</tr>
<tr>
<td></td>
<td>2275.78</td>
<td></td>
<td>275.85</td>
<td></td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>pure breathy</td>
<td>3.53573</td>
<td>1.22650</td>
<td>0.061</td>
</tr>
<tr>
<td></td>
<td>3.28384</td>
<td></td>
<td>2.17074</td>
<td></td>
</tr>
</tbody>
</table>

Table 6.8.2 Means and standard deviations of the glottal source and formant parameters for Gujarati female /e/, averaged across the three tokens of the vowel as well as the initial and mid measures of the first subsegment. The results of the one-way ANOVA are given in the last two columns.
vowel. As $-3.951$ dB/octave is substantially steeper than the above values, it almost certainly represents lax voice.

The mid vowel /e/, on the other hand, shows a wide spread between the mean HNRs of breathy voice ($9.533$ dB) and pure voice ($19.316$ dB). There are also significant differences between the spectral tilts (harmonic slope, H1–H2) of the two phonation types. Recall, however, that when a noisy phonation type occurs, i.e. when there is a large amount of either aspiration noise (low HNR: *breathy voice*) or modulation noise (low log pACC: *harsh voice*) or both (low HNR and low log pACC: *harsh whispery voice*), the harmonic slope diminishes in significance and becomes a minor acoustic cue (see Sections 2.4.1 and especially 2.4.2.4).

In Table B.1.9 the HNRs of the phonologically pure voice female vowel /e/ in mel ‘dirt’ display exceptionally low values: $9.783$ dB (initial measure), $12.794$ dB (mid measure). For this reason, the vowel would normally be considered breathy since the HNRs lie well below $15-18$ dB, the most likely upper threshold range for aspiration noise (see Section 5.4.2). On the other hand, in Table B.1.10 the HNRs of the breathy voice vowel /e/ in mel ‘palace’ show even lower values, $8.278$ dB (initial measure), $3.763$ dB (mid measure), when the respective time points are compared pairwise. Thus the phonological opposition pure voice vs. breathy voice is maintained despite the fact that the absolute HNRs seem to indicate its neutralization. For a given speaker, it is therefore important to compare only measures occurring in a similar phonetic context defined in terms of temporal location and motor-acoustic properties (cf. also the discussion of Fant, 1973: 210 in Section 4.3). Statistical averages are useful indicators of overall trends, but care must be exercised in their interpretation.
<table>
<thead>
<tr>
<th>Gujarati: man</th>
<th>/a/</th>
<th>Mean</th>
<th>S.D.</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>F0 (Hz)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pure</td>
<td></td>
<td>157.15</td>
<td>13.54</td>
<td>0.59</td>
<td>0.814</td>
</tr>
<tr>
<td>breathy</td>
<td></td>
<td>159.03</td>
<td>13.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pure</td>
<td></td>
<td>-1.8077</td>
<td>1.3115</td>
<td>1.406</td>
<td>0.263</td>
</tr>
<tr>
<td>breathy</td>
<td></td>
<td>-1.1306</td>
<td>0.4857</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Log pACC (dB)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pure</td>
<td></td>
<td>23.330</td>
<td>0.000</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>breathy</td>
<td></td>
<td>23.330</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pure</td>
<td></td>
<td>-3.544</td>
<td>0.275</td>
<td>2.173</td>
<td>0.171</td>
</tr>
<tr>
<td>breathy</td>
<td></td>
<td>-3.889</td>
<td>0.503</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>HNR (dB)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pure</td>
<td></td>
<td>23.330</td>
<td>0.000</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>breathy</td>
<td></td>
<td>23.330</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Slope (dB/octave)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pure</td>
<td></td>
<td>2.217</td>
<td>2.085</td>
<td>3.944</td>
<td>0.075</td>
</tr>
<tr>
<td>breathy</td>
<td></td>
<td>5.144</td>
<td>2.948</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>H1–H2 (dB)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pure</td>
<td></td>
<td>577.73</td>
<td>87.49</td>
<td>3.600</td>
<td>0.087</td>
</tr>
<tr>
<td>breathy</td>
<td></td>
<td>684.95</td>
<td>107.27</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>F1 (Hz)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pure</td>
<td></td>
<td>5.12625</td>
<td>1.30864</td>
<td>0.513</td>
<td>0.490</td>
</tr>
<tr>
<td>breathy</td>
<td></td>
<td>5.75017</td>
<td>1.68601</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PE1 (dB)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pure</td>
<td></td>
<td>1103.28</td>
<td>154.51</td>
<td>6.409</td>
<td>0.030</td>
</tr>
<tr>
<td>breathy</td>
<td></td>
<td>1300.22</td>
<td>111.53</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>PE2 (dB)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pure</td>
<td></td>
<td>3.91576</td>
<td>0.64913</td>
<td>2.547</td>
<td>0.142</td>
</tr>
<tr>
<td>breathy</td>
<td></td>
<td>2.65419</td>
<td>1.82421</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6.9.1 Means and standard deviations of the glottal source and formant parameters for Gujarati male /a/, averaged across the three tokens of the vowel as well as the initial and mid measures of the first subsegment. The results of the one-way ANOVA are given in the last two columns.
Unlike the woman’s vowels, neither the HNR nor the spectral tilt distinguish male
pure and breathy vowels in a consistent manner. Although the means of the HNR and
spectral tilt lie in the correct direction for both /a/ and /e/, the narrow-band slope H1–H2
of male /e/ is the sole vocal source measure to achieve a significance level below 0.05.
Similarly to the woman’s analysis, /e/ is the only male vowel to show a contrast in HNR.

In Table B.2.10 the HNR of the initial measure of the breathy ə in məl ‘palace’ is 12.794
dB. All other pure and breathy HNRs exhibit the maximum value of 23.330 dB, including
the initial measure of the pure vowel in məl ‘dirt’ in Table B.2.9. As pointed out at the
beginning of this section, contrastive breathy voice has a marginal phonological status in

<table>
<thead>
<tr>
<th>Gujarati: man</th>
<th>/e/</th>
<th>Mean</th>
<th>S.D.</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>F0 (Hz)</td>
<td>pure</td>
<td>155.59</td>
<td>12.31</td>
<td>0.004</td>
<td>0.954</td>
</tr>
<tr>
<td></td>
<td>breathy</td>
<td>156.03</td>
<td>13.73</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>pure</td>
<td>-1.2647</td>
<td>0.8639</td>
<td>0.397</td>
<td>0.543</td>
</tr>
<tr>
<td></td>
<td>breathy</td>
<td>-1.0384</td>
<td>0.1650</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>pure</td>
<td>23.330</td>
<td>0.000</td>
<td>1.000</td>
<td>0.341</td>
</tr>
<tr>
<td></td>
<td>breathy</td>
<td>21.574</td>
<td>4.301</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>pure</td>
<td>-3.873</td>
<td>0.307</td>
<td>1.320</td>
<td>0.277</td>
</tr>
<tr>
<td></td>
<td>breathy</td>
<td>-4.476</td>
<td>1.250</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>pure</td>
<td>-1.247</td>
<td>3.797</td>
<td>17.115</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>breathy</td>
<td>6.126</td>
<td>2.155</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>pure</td>
<td>543.08</td>
<td>77.84</td>
<td>0.362</td>
<td>0.561</td>
</tr>
<tr>
<td></td>
<td>breathy</td>
<td>569.61</td>
<td>74.78</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>pure</td>
<td>6.12013</td>
<td>2.18373</td>
<td>0.009</td>
<td>0.925</td>
</tr>
<tr>
<td></td>
<td>breathy</td>
<td>6.02491</td>
<td>1.03240</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>pure</td>
<td>1921.23</td>
<td>98.47</td>
<td>0.725</td>
<td>0.415</td>
</tr>
<tr>
<td></td>
<td>breathy</td>
<td>1879.12</td>
<td>70.59</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>pure</td>
<td>3.81672</td>
<td>0.99776</td>
<td>0.480</td>
<td>0.504</td>
</tr>
<tr>
<td></td>
<td>breathy</td>
<td>4.25041</td>
<td>1.16484</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 6.9.2 Means and standard deviations of the glottal source and formant
parameters for Gujarati male /e/, averaged across the three tokens of the vowel as
well as the initial and mid measures of the first subsegment. The results of the
one-way ANOVA are given in the last two columns.

Unlike the woman’s vowels, neither the HNR nor the spectral tilt distinguish male
pure and breathy vowels in a consistent manner. Although the means of the HNR and
spectral tilt lie in the correct direction for both /a/ and /e/, the narrow-band slope H1–H2
of male /e/ is the sole vocal source measure to achieve a significance level below 0.05.
Similarly to the woman’s analysis, /e/ is the only male vowel to show a contrast in HNR.

In Table B.2.10 the HNR of the initial measure of the breathy ə in məl ‘palace’ is 12.794
dB. All other pure and breathy HNRs exhibit the maximum value of 23.330 dB, including
the initial measure of the pure vowel in məl ‘dirt’ in Table B.2.9. As pointed out at the
beginning of this section, contrastive breathy voice has a marginal phonological status in
Gujarati. Hence it is not surprising that an unusually large amount of idiolectal variation exists.

To conclude this section on Gujarati, observe that the female mid vowel /e/, which unambiguously contrasts pure and breathy voice, shows a higher average PE1 for breathy voice (6.25517 dB) than it does for pure voice (5.62854 dB). This means that the PE1 is a very poor cue for the glottal opening of voiced sounds. Contrary to what occurs for the glottal noise sounds glottal stop, whisper, breath, the difference in glottal area between pure voice and breathy voice is apparently not sufficient to be reflected in this acoustic parameter.

### 6.3.2 Jalapa Mazatec

The Otomanguean language Jalapa Mazatec displays a three-way contrast between phonologically pure, harsh, and breathy voice (Kirk et al., 1993; Silverman et al., 1995). One male and one female speaker of the language produced six isolated words extracted from a much more comprehensive field recording (see Blankenship, 1997 for additional references and the recording circumstances.). The two pure voice words are sa³ ‘moon’, sa² ‘he sings’; the two harsh voice words are tʃa³ ‘load, burden’, ŋdæ¹ ‘buttocks’; the two breathy voice words are ŋdɛ̃³ ‘hard’, ŋdɛ̄¹ ‘horse’. The tone marks are Low (1), Mid (2), and High (3). The female glottal source and formant measures of the six words are shown in Appendix C (Tables C.1.1-C.1.6); the male measures are likewise given in Appendix C (Tables C.2.1-C.2.6). If there is more than one subsegment, an asterisk (*) indicates the one with the maximum energy.
To increase the number of data points available, the pure, harsh, and breathy measures are averaged across the two low vowels /a/ and /æ/ as well as the initial and mid frames of the first subsegment, yielding four samples of each phonation type. For each Jalapa Mazatec speaker, two independent one-way ANOVAs are applied, the first on pure voice (N = 4) and harsh voice (N = 4), the second on pure voice (N = 4) and breathy voice (N = 4). The woman’s results are shown in Tables 6.10.1-6.10.2, the man’s in Tables 6.11.1-6.11.2.

<table>
<thead>
<tr>
<th>Mazatec: woman</th>
<th>Type</th>
<th>Mean</th>
<th>S.D.</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>F0 (Hz)</td>
<td>pure</td>
<td>229.60</td>
<td>30.55</td>
<td>0.885</td>
<td>0.383</td>
</tr>
<tr>
<td></td>
<td>harsh</td>
<td>195.70</td>
<td>65.29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>pure</td>
<td>–0.6193</td>
<td>0.1241</td>
<td>0.885</td>
<td>0.383</td>
</tr>
<tr>
<td></td>
<td>harsh</td>
<td>–1.9312</td>
<td>0.6094</td>
<td>17.800</td>
<td>0.006</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>pure</td>
<td>23.330</td>
<td>0.0000</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td></td>
<td>harsh</td>
<td>23.330</td>
<td>0.0000</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>pure</td>
<td>–4.656</td>
<td>0.455</td>
<td>7.943</td>
<td>0.030</td>
</tr>
<tr>
<td></td>
<td>harsh</td>
<td>–3.881</td>
<td>0.311</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>pure</td>
<td>–4.668</td>
<td>3.316</td>
<td>0.353</td>
<td>0.574</td>
</tr>
<tr>
<td></td>
<td>harsh</td>
<td>–6.242</td>
<td>4.138</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>pure</td>
<td>564.50</td>
<td>56.34</td>
<td>0.359</td>
<td>0.571</td>
</tr>
<tr>
<td></td>
<td>harsh</td>
<td>601.04</td>
<td>108.18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>pure</td>
<td>6.18190</td>
<td>1.57395</td>
<td>0.696</td>
<td>0.436</td>
</tr>
<tr>
<td></td>
<td>harsh</td>
<td>5.25056</td>
<td>1.58419</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>pure</td>
<td>1734.31</td>
<td>238.05</td>
<td>1.173</td>
<td>0.320</td>
</tr>
<tr>
<td></td>
<td>harsh</td>
<td>1873.12</td>
<td>95.053</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>pure</td>
<td>2.28526</td>
<td>1.58078</td>
<td>0.583</td>
<td>0.474</td>
</tr>
<tr>
<td></td>
<td>harsh</td>
<td>3.17987</td>
<td>1.72849</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 6.10.1** Means and standard deviations of the glottal source and formant parameters of Jalapa Mazatec female pure and harsh measures, averaged across the two low vowels /a/ and /æ/ as well as the initial and mid points of the first subsegment. The results of the one-way ANOVA are given in the last two columns.
Among the woman’s vocal source parameters in Table 6.10.1, the log pACC displays the most significant difference between the means of pure voice (–0.6193 dB) and harsh voice (–1.9312 dB), as would be expected if this parameter is the major cue for modulation noise. In Table 5.5 the log pACC of synthesized harsh voice at 229.72 Hz is –1.11 dB, a value substantially higher than –1.9312 dB at the mean F0 of 195.70 Hz. Thus there is considerable modulation noise during the first portion of the woman’s vowels /a/ and /æ/. The highest possible HNR of 23.330 dB shows that even a large amount of modulation noise has little influence on the HNR in actual speech, which constitutes further proof of the independence of the proposed HNR measure vis-à-vis

### Table 6.10.2

Means and standard deviations of the glottal source and formant parameters of Jalapa Mazatec female pure and breathy measures, averaged across the two low vowels /a/ and /æ/ as well as the initial and mid points of the first subsegment. The results of the one-way ANOVA are given in the last two columns.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Type</th>
<th>Mean</th>
<th>S.D.</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>F0 (Hz)</td>
<td>pure</td>
<td>229.60</td>
<td>30.55</td>
<td>0.678</td>
<td>0.442</td>
</tr>
<tr>
<td></td>
<td>breathy</td>
<td>213.87</td>
<td>22.97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>pure</td>
<td>–0.6193</td>
<td>0.1241</td>
<td>0.269</td>
<td>0.623</td>
</tr>
<tr>
<td></td>
<td>breathy</td>
<td>–0.6614</td>
<td>0.1046</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>pure</td>
<td>23.330</td>
<td>0.000</td>
<td>1.000</td>
<td>0.356</td>
</tr>
<tr>
<td></td>
<td>breathy</td>
<td>21.449</td>
<td>3.763</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>pure</td>
<td>–4.656</td>
<td>0.455</td>
<td>1.783</td>
<td>0.230</td>
</tr>
<tr>
<td></td>
<td>breathy</td>
<td>–5.503</td>
<td>1.184</td>
<td></td>
<td></td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>pure</td>
<td>–4.668</td>
<td>3.316</td>
<td>8.833</td>
<td>0.025</td>
</tr>
<tr>
<td></td>
<td>breathy</td>
<td>2.595</td>
<td>3.589</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>pure</td>
<td>564.50</td>
<td>56.34</td>
<td>3.250</td>
<td>0.122</td>
</tr>
<tr>
<td></td>
<td>breathy</td>
<td>642.99</td>
<td>66.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>pure</td>
<td>6.18190</td>
<td>1.57395</td>
<td>1.575</td>
<td>0.256</td>
</tr>
<tr>
<td></td>
<td>breathy</td>
<td>4.92289</td>
<td>1.24407</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>pure</td>
<td>1734.31</td>
<td>238.05</td>
<td>0.026</td>
<td>0.878</td>
</tr>
<tr>
<td></td>
<td>breathy</td>
<td>1691.25</td>
<td>480.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>pure</td>
<td>2.28526</td>
<td>1.58078</td>
<td>0.117</td>
<td>0.744</td>
</tr>
<tr>
<td></td>
<td>breathy</td>
<td>2.55969</td>
<td>0.27204</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
modulation noise, previously demonstrated in Chapter 5 for synthetic glottal waveforms. For the harmonic slope, the difference between the means of pure and harsh voice is also significant (Sig. < 0.05), but since the modulation noise is very strong, the harmonic slope serves only as a minor acoustic cue.

A comparison of woman’s pure and breathy means in Table 6.10.2 indicates that the spectral tilt is the relevant parameter, as the HNRs of pure and breathy voice are almost equally high and the difference between the means of pure and breathy voice is significant for the narrow-band spectral tilt H1–H2 (Sig. = 0.025). Thus the pure vs. breathy phonological opposition is realized as a contrast of plain and lax voice. Note that the difference between the means of the wide-band harmonic slope is not significant (Sig. = 0.230), although the average values of −4.656 dB/octave (pure voice) and −5.503 dB/octave (breathy voice) show that the harmonic slope of phonologically breathy voice is steeper.

In the course of Section 6.2, it was pointed out that the narrow-band spectral tilt H1–H2 not only performs poorly as an indicator of spectral tilt in Mpi, but there is also an important theoretical consideration against it: the glottal bandwidth (GBW) put forward at the end of Chapter 2 (Section 2.4.2.4) corresponds to the effective bandwidth over the entire speech spectrum, suggesting that the wide-band harmonic slope is a more appropriate measure of spectral tilt than the narrow-band slope H1–H2. Yet the narrow-band spectral tilt H1–H2 performs better than the wide-band harmonic slope in distinguishing between phonologically breathy and pure voice for the female speaker of Jalapa Mazatec as well as for the male speaker of Gujarati discussed earlier (e.g. the vowel /e/). One way of resolving this apparent contradiction is to recall from Chapter 3
that in the time domain the wide-band spectral tilt, the harmonic slope in this case, is primarily an indicator of the abruptness of the glottal pulse whereas the narrow-band spectral tilt H1–H2 appears to be largely dependent on the OQ (open quotient) and the symmetry of the glottal pulse. The differential sensitivity of the narrow- and wide-band spectral tilts to specific glottal waveform patterns may therefore reflect the tendency of the auditory speech mechanism to provide redundancy against environmental noise and perceptual errors by using two information channels. In view of the phonological significance of the broad-spectrum GBW as a means of ordering the nine phonation types along a linear scale, it nevertheless remains likely that the wide-band harmonic slope tilt is the major cue for spectral tilt.
Table 6.1.1 Means and standard deviations of the glottal source and formant parameters of Jalapa Mazatec male pure and harsh measures, averaged across the two low vowels /a/ and /æ/ as well as the initial and mid points of the first subsegment. The results of the one-way ANOVA are given in the last two columns.
For the male Jalapa Mazatec speaker, the harmonic slope is the only significant vocal source parameter distinguishing pure and harsh voice (Sig. = 0.048) while \( H1–H2 \) is the only significant parameter separating pure voice and breathy voice (Sig. = 0.017). Because the distinction is one of spectral tilt in both cases, there appears to be a three-way phonemic contrast: tense voice (–3.599 dB/octave) vs. plain voice (–4.116 dB/octave) vs. lax voice (–5.597 dB/octave). Remark that the harmonic slope of tense voice is nearer to values more typical of plain voice in Table 5.12 (natural male \( \alpha \): –3.49 dB/octave, synthetic male \( \alpha \): –3.53 dB/octave, natural male \( \alpha \) : –3.60 dB/octave, synthetic male \( \alpha \) : –4.00 dB/octave). The fairly steep harmonic slope of plain voice parallels the

<table>
<thead>
<tr>
<th>Mazatec: man</th>
<th>Type</th>
<th>Mean</th>
<th>S.D.</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>pure</td>
<td>191.82</td>
<td>41.93</td>
<td>0.125</td>
<td>0.736</td>
</tr>
<tr>
<td></td>
<td>harsh</td>
<td>181.65</td>
<td>39.51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>pure</td>
<td>−0.8099</td>
<td>0.2560</td>
<td>1.680</td>
<td>0.243</td>
</tr>
<tr>
<td></td>
<td>harsh</td>
<td>−1.6066</td>
<td>1.2023</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>pure</td>
<td>23.330</td>
<td>0.000</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td></td>
<td>harsh</td>
<td>23.330</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>pure</td>
<td>−4.116</td>
<td>0.371</td>
<td>6.101</td>
<td>0.048</td>
</tr>
<tr>
<td></td>
<td>−3.599</td>
<td>0.194</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>pure</td>
<td>−1.235</td>
<td>2.293</td>
<td>0.473</td>
<td>0.517</td>
</tr>
<tr>
<td></td>
<td>−2.315</td>
<td>2.146</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>pure</td>
<td>655.42</td>
<td>140.95</td>
<td>1.447</td>
<td>0.274</td>
</tr>
<tr>
<td></td>
<td>746.95</td>
<td>57.40</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>pure</td>
<td>5.98389</td>
<td>1.48360</td>
<td>0.330</td>
<td>0.586</td>
</tr>
<tr>
<td></td>
<td>6.71088</td>
<td>2.04942</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>pure</td>
<td>1661.14</td>
<td>226.87</td>
<td>0.010</td>
<td>0.923</td>
</tr>
<tr>
<td></td>
<td>1642.12</td>
<td>300.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>pure</td>
<td>3.63546</td>
<td>1.25764</td>
<td>0.050</td>
<td>0.831</td>
</tr>
<tr>
<td></td>
<td>3.82031</td>
<td>1.07579</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 6.11.2** Means and standard deviations of the glottal source and formant parameters of Jalapa Mazatec male pure and breathy measures, averaged across the two low vowels /a/ and /æ/ as well as the initial and mid points of the first subsegment. The results of the one-way ANOVA are given in the last two columns.

For the male Jalapa Mazatec speaker, the harmonic slope is the only significant vocal source parameter distinguishing pure and harsh voice (Sig. = 0.048) while \( H1–H2 \) is the only significant parameter separating pure voice and breathy voice (Sig. = 0.017).
very low HNR of the nominally pure voice of female Gujarati /e/ in Table B.1.9. Despite recurring cross-linguistic trends, these data make it clear that the acoustic continuum is divided phonologically in a language-specific way.

As was pointed out in the introduction to Chapter 2, *harsh whispery voice* never appears to enter into lexically contrastive oppositions according to available phonological descriptions. Laver (1980: 131-132) refers to it parenthetically when he mentions that some urban Scots accents are characterized by ventricular or possibly whispery ventricular voice. However neither *harsh whispery voice* nor any other variant phonation type with co-occurring modulation and aspiration noise is among the descriptors Esling (1978) uses in his phonetically oriented sociolinguistic study of Edinburgh speech. In addition to an apparent lack of lexical contrastiveness, another reason for the neglect of *harsh whispery voice* may stem from the fact that modulation and aspiration noise are often not distinguishable in spectrographic displays of speech (cf. Hillenbrand, 1987). As is evident from the literature review of Chapter 3, the search for independent measures of modulation and aspiration noise gained momentum only at the beginning of the last decade.

Inspection of the final measures of the last vocalic subsegment, if there is more than one, reveals a number of instances of *harsh whispery voice*. For the female Mazatec speaker, the final measures of nonpure (breathy and harsh) voice have HNRs and log pACCs below 12 dB and −1.55 dB respectively (Tables C.1.3-C.1.6). Because the F0s are all above 165 Hz, a log pACC of less than −1.55 dB is on the same order as the synthesized harsh glottal waves in Table 5.5 (log pACC = −2.45 at 73.27 Hz;
log $p_{\text{ACC}} = -1.46$ dB at 129.29 Hz; log $p_{\text{ACC}} = -1.11$ dB at 229.93 Hz). A HNR of 12 dB lies below the assumed upper threshold range of 15-18 dB for aspiration noise.

The male Mazatec speaker shows a similar tendency to employ harsh whispery voice in final position (phonologically pure voice in Table C.2.2: HNR = 9.783 dB, log $p_{\text{ACC}} = -2.4203$ dB; phonologically harsh voice in Table C.2.3: HNR = 8.278 dB, log $p_{\text{ACC}} = -2.4893$ dB; phonologically breathy voice in Table C.2.5: HNR = 12.794 dB, log $p_{\text{ACC}} = -2.8136$ dB). There is also one case in which the male Mpi speaker produced final harsh whispery voice (tense low level /si/ in Table A.10: F0 = 70.81 Hz, HNR = 11.289 dB, log $p_{\text{ACC}} = -6.1659$ dB).

These data demonstrate that harsh whispery voice is far from being a rare phenomenon. In terms of phonemic distinctiveness, the examples illustrating this phonation type are strongly neutralizing since harsh whispery voice is coextensive with the final time sample of phonologically pure, harsh, or breathy vowels. Because there seems to be a slight tendency for final harsh whispery voice to occur preferentially with phonologically harsh vowels, it may represent the combined effect of the constricted glottis of harsh voice and the spread glottis of breath, as final devoicing is not infrequent cross-linguistically. Note that the surprisingly high frequency of occurrence of harsh whispery voice shows the great importance of keeping modulation and aspiration noise distinct in the acoustic assessment of phonation type.

6.4 General summary of the phonetic analysis of the phonation types

In this chapter several languages with known phonemic contrasts in phonation type are investigated. The speech analysis procedures developed in Chapter 4 appear to
distinguish rather well among the nine phonation types, each of which is represented by at least one particularly good exemplar. Of the glottal noise phonation types, a phonological opposition between the glottal stop /t/ and voicelessness /h/ (whisper or breath) is the most common one. A glottal stop is produced with a sealed glottis during its inaudible closed phase and a slightly open glottis during the audible transient of its attack or release phase. First formant damping caused by shunt losses is negligible when the glottis is constricted during the glottal stop, but the attenuation progressively increases with the larger glottal opening of whisper, and even more so with the spread glottis of breath. An increase in glottal damping leads to a broadening mainly of F1 bandwidth (B1) and a corresponding reduction in the F1 peak energy factor (PE1).

When the initial, mid, and final time frames of postvocalic /t/ and /h/ are compared pairwise in Avar, the PE1s are all higher for /t/ than they are for /h/, in conformity with the anticipated results (Table 6.2).

Quechua has a three-way contrast between voiceless, aspirated voiceless, and ejective obstruents. As the ejective is generated by an initiatory glottal stop, its supraglottal release should be accompanied by a larger PE1 value than those of the voiceless and aspirated voiceless stops. As expected, the uvular and velar ejectives (Table 6.4.1 and Table 6.4.2, respectively) as well as the fricative portion of the palato-alveolar affricate ejective (Table 6.4.4) exhibit higher PE1s than the corresponding voiceless and aspirated voiceless obstruents at the mid measure, the one with the presumed greatest robustness against coarticulatory effects. The unaspirated velar and the fricative portion of the unaspirated palato-alveolar affricate likewise have higher PE1s than the aspirated ones at the midpoint of the release of the obstruent, very probably indicating a three-way
contrast of increasing glottal opening at these two places of articulation: *glottal stop*, *whisper* and *breath*. The fact that the uvular series only opposes the glottal stop to voicelessness (*whisper* or *breath*) while the distinction between *whisper* or *breath* is neutralized also manifests itself in the world’s languages, the phonologies of which rarely show evidence of a contrast between these two kinds of voicelessness (see the discussion on Proto-Tai tones in Section 2.4.1 of Chapter 2).

Pure voice is defined as voiced phonation without significant aspiration or modulation noise. Since the likely upper HNR threshold of aspiration noise is 15-18 dB, the HNR should lie above this range in order for voiced phonation to be considered phonetically pure voice (Section 5.4.2). Similarly, the log pACC of pure voice at a given fundamental frequency should approach the values of synthetic normal voice and considerably exceed those of synthetic harsh voice provided in Table 5.5. Tense voice is defined as pure voice with a spectral tilt flatter than plain voice, plain voice as pure voice with a spectral tilt lying between lax voice and tense voice, and lax voice as pure voice with a spectral tilt steeper than plain voice.

The phonological opposition between plain voice and tense voice is illustrated by Mpi. The mean harmonic slope of plain /i/ is much closer to the values of the natural and synthesized vowels in Table 5.12 than the flatter slope of tense /i/. In Gujarati, the phonological opposition between female pure and breathy /a/ is realized as a phonetic contrast between plain and lax voice, with the corresponding significant difference in mean harmonic slope (plain voice: –3.253 dB/octave, lax voice: –3.951 dB/octave).

Although the narrow-band spectral tilt H1–H2 may often yield better results than the wide-band harmonic slope in distinguishing among the three pure voice phonation
types, the wide-band harmonic slope is considered the major cue for spectral tilt in view of the importance of the broad-spectrum GBW as a means of ordering the nine phonation types along a linear scale (see Section 2.4.2.4 in Chapter 2). The narrow-band spectral tilt H1–H2 may be an additional information channel serving as a redundant backup against environmental noise and perceptual errors.

Contrary to pure voice, noisy voice is defined as voiced phonation with a large amount of aspiration noise or modulation noise. Breathy voice is associated with significant aspiration noise, harsh (or laryngealized) voice with significant modulation noise, and harsh whispery voice with both kinds of noise.

Gujarati has a marginal phonemic contrast between pure and breathy vowels. Good examples of the distinctive function of aspiration noise are provided by two minimal pairs of the woman speaker: 1) pelo ‘that one’ vs. pêlo ‘first’ (Tables B.1.7-B.1.8) and 2) wer ‘revenge, animosity’ vs. wêr ‘sawdust’ (Tables B.1.11-B.1.12). The initial, mid, and final HNR measures of pelo are respectively 23.330 dB, 23.330 dB, 23.330 dB, those of pêlo are 11.289 dB, 11.289 dB, 11.289 dB; the initial, mid, and final HNR measures of wer are respectively 23.330 dB, 23.330 dB, 23.330 dB, those of wêr are 8.278 dB, 14.299 dB, 23.330 dB.

Jalapa Mazatec has a three-way opposition between pure, harsh, and breathy vowels. A good illustration of the distinctive function of modulation noise is given by the near minimal pair of the woman speaker: sa³ ‘moon’ vs. tʃa³ ‘load, burden’ (Table C.1.1 and Table C.1.3). The initial, mid, and final log pACC measures of sa³ are respectively –0.5649 dB (257.86 Hz), –0.5569 dB (254.16 Hz), –0.9232 dB (243.38 Hz), those of the
first subsegment of $t\hat{a}$ are $-1.4521$ dB (254.16 Hz), $-1.7640$ dB (239.89 Hz) $-3.7929$ dB (207.64 Hz).

Both the man’s and the woman’s final measures of Jalapa Mazatec show a number of examples of harsh whispery voice, in which large amounts of both aspiration noise (low HNR) and modulation noise (low log pACC) are present. In addition, the male Mpi speaker produced one token of harsh whispery voice in final position (tense low level /si/ in Table A.10). These results indicate that harsh whispery voice occurs much more frequently than previous work on the phonetics and phonology of phonation type would suggest. One possible reason for the neglect of harsh whispery voice is that traditional spectral displays do not clearly distinguish between aspiration and modulation noise.

Now that the phonetics of the nine phonation types has been placed on somewhat firmer ground, the next chapter will return to the phonological issues raised at the end of Chapter 2.
NOTES TO CHAPTER 6

1. The URL of the Thai speech samples:

   http://web.uvic.ca/ling/resources/ipa/handbook/Thai.zip

   Note that the beginning and end points of the subsegmental duration measures in Table 6.1 do not correspond exactly to the time scale of the spectrographic display in Figure 6.1 (e.g. 139.7 ms and 147.1 ms for the release burst of ?) since the 39.37 ms rectangular window is longer than the effective window length of the wide-band spectrogram. This holds for the other spectrograms as well.

2. The URL of the Quechua speech samples:

   http://www.phonetics.ucla.edu/vowels/chapter12/quechua.html

3. The URL of the Mpi speech samples:

   http://www.phonetics.ucla.edu/vowels/chapter12/mpi.html

4. All the Gujarati informants were paid through a Fred W. Householder Memorial research grant awarded by the Indiana University Linguistics Department.

5. Daniel Silverman digitized the Jalapa Mazatec recordings referenced in Blankenship (1997) and graciously sent me the audio CDs. Each of the words was spoken outdoors by a man and then by five women in succession. The speech samples consist of those produced by the man and the first woman to speak.
CHAPTER 7
GLOTTAL MANNER FEATURES

7.1 An overview of glottal manner feature systems

7.1.1 Introduction

The objective of this section is to introduce the various classifications of phonation type in chronological order. The most influential feature systems of glottal manner are therefore described and critically reviewed. Particular emphasis is given at the beginning of the section to Trubetzkoy’s account of glottal manner because—similarly to his—all later systems take the possibly complex phonemic segment as the basic unit of feature analysis rather than the phonetically homogeneous subsegment. This practice makes it difficult to establish a correspondence between a given phonological feature and its acoustic manifestation in the speech waveform. A further point considered is that an association between features should not be built permanently into the feature system, but rather expressed as a co-occurrence rule in the individual language or as a typological tendency. The final part of this section is devoted to privative features of glottal manner and a discussion of their phonetic and phonological shortcomings.

7.1.2 Trubetzkoy (1938) and the complex segment

Trubetzkoy was about twenty pages short of completing his *Grundzüge der Phonologie* (1967) before his untimely death in 1938. The manuscript thus represents his last contribution to the study of phonology. The opposition voiceless vs. voiced is subsumed under his obstruent manner properties (*Überwindungseigenschaften*), which
together with place properties (Lokalisierungseigenschaften) and resonance properties such as nasality (Resonanz eigenschaften) constitute the consonant properties (konsonantische Eigenschaften, 1967: 114-166). Not only is the opposition voiceless vs. voiced considered a type of obstruent manner correlation of the second degree (Überwindungsartkorrelation zweites Grades), but the oppositions aspirated vs. unaspirated and plosive vs. ejective are so as well, among others (1967: 140).

In the survey of phonation types of Chapter 2, it was pointed out that Ladefoged (1971) defined aspiration as a brief period of voicelessness during and immediately after the release of the articulatory stricture. Furthermore, it was mentioned that unlike the other phonation types, aspiration lies outside his glottal stricture ranking since it also involves timing considerations. Later he and Maddieson (1996: 70) further refined the definition to be as follows:

“…aspiration is a period after the release of a stricture and before the start of regular voicing (or the start of another segment, or the completion of an utterance) in which the vocal folds are markedly further apart than they are in modally voiced sounds.”

The definition takes into account the fact that aspirated voiceless obstruents often exhibit a wider glottis at release and have a longer voice onset time (VOT) than unaspirated voiceless obstruents. For example, Kim’s cineradiographic study (1970) of Korean shows that the aspirated stops have a considerably wider glottis at release than either the lenis (slightly aspirated) stops or the fortis (unaspirated) stops, with the former displaying a somewhat wider glottis than the latter (see also Löfqvist and Yoshioka, 1981). The average VOT of the aspirated series is 90 ms, the lenis series 35 ms, and the unaspirated series 10 ms. These results generally concord with the Quechua acoustical analysis in Chapter 6, in which the velar and the palato-alveolar affricate manifest a three-way
contrast of increasing glottal opening at the release of the articulatory stricture, with the phonologically aspirated sound being the widest (breath), the phonologically unaspirated sound the next widest (whisper), and the ejective the narrowest (glottal stop). On the other hand, although the Quechua uvular series maintains the opposition between the glottal stop (ejective) and voicelessness (plosives), the distinction between aspirated plosive (breath voicelessness) and unaspirated plosive (whisper voicelessness) appears to be neutralized (cf. Table 6.4.1). Because aspirated voiceless obstruents may not always be implemented with the most open glottis at articulatory release, aspiration is not universally realized with a widely spread glottis, as proposed by Iverson and Salmons (1995).

Trubetzkoy’s classification clearly indicates that he considers the phonological relation between aspirated and unaspirated stops to be parallel to the one between voiceless and voiced stops. An excessive emphasis on formal phonological relations, however, can obscure significant differences in phonetic substance. One manifestation of this problem is the attempt to assign a unitary set of features to a segment that functions as one phonologically, yet consists of a sequence of distinct phonetically homogeneous subsegments. Furthermore, designations such as the traditional “aspirated p” or even “p realized with a long VOT” reinforce the impression that the bilabial plosive has undergone a concurrent featural modification like voicing rather than being followed in time by another subsegment.

Detailed analyses of intraphonemic temporal relations conducted over the past two decades have revealed that doubly-articulated stops like /kp, qb/ differ from simple plosives in that they consist of a time-ordered sequence of movements by the two
articulators (see Ladefoged and Maddieson, 1996: 332-343 and references therein).

Hence it is of all the more interest to consider briefly the traditional and current ways in which subsegments are represented.

The phonemic principle is stated in Section 4 of the *Handbook of the International Phonetic Association* (1999: 27):

“In its earliest days...the International Phonetic Association has aimed to provide ‘a separate sign for each distinctive sound; that is, for each sound which, being used instead of another, in the same language, can change the meaning of a word.’ This notion of a ‘distinctive sound’ is what became widely known in the twentieth century as the phoneme. Its history is far longer, though. For instance, the phonemic principle is implicit in the invention of alphabet writing. However a lot of languages, such as English, have spelling systems in which the relation between phonemes and letters of the alphabet has become obscured. This very fact was a motivation for the creation of a universally agreed system of phonetic notation. So in English, the IPA provides a symbol /k/ which stands unambiguously for the phoneme which is variously written as <c> (*car*), <k> (*kettle*), <ck> (*back*), <ch> (*monarch*), <q> (*quick*), and in other ways.

...Conventionally, as in the English example above, symbols for the phonemes of the language are placed within oblique lines / /.

In general, the symbol for a phoneme will be an unmodified letter of the IPA, but letters may also be combined to make a phoneme symbol (for instance /tʃ/, as at the beginning and end of English *church*; if necessary the phonological unity of the two segments can be shown by a tie bar: /tʃ/).

The tie bar diacritic is therefore used to indicate that the subsegmental stop and following fricative belong to one affricate phoneme. Note that the left-to-right spatial order is an iconic cue to the temporal order of the subsegments, as is typically the case in European writing systems. A more frequent indication of subsegmental constituency is the use of superscripts, most often again in the left-to-right iconic order, such as with the affricate /tʃ/ (= /tʃ/), or the prenasalized /n̪t/ (= /n̪t/) and postnasalized stops /n̩t̪/ (= /n̩t̪/).

However the left-to-right ordering convention does not seem to hold for secondary articulations like palatalization, (labio)velarization, and pharyngealization, in which the
right superscript \(^j\), for instance, is usually taken to mean tongue body fronting throughout the entire segmental phoneme (/t\(^j\) ≠ /t\(^j\)/). To avoid misunderstanding, the usage of the superscript should be explicitly defined on the basis of acoustic and, ideally, physiological measures in the language concerned.

An approach very similar to the IPA representation of complex segments is developed by St. Clair (1972, 1973; cf. also Hoard, 1967) who proposed that the subsegmental constituents of a phoneme be bracketed, instead of being grouped together by a tie bar, through the Compound Segment Convention: \([X] [Y] \rightarrow [[X][Y]]\). For instance, when this convention is applied to the segments [t] and [ʃ], it yields the affricate [[t][ʃ]]. He points out that (1972):

“By means of this convention we are also able to characterize the fact noted by Jakobson that languages with affricates also have dental stops and palatal sibilants. That is to say, languages in which [č] also have [t] and [š]. The former presupposes the latter two. Since a compound phoneme has an inner structure, we adequately incorporate Jakobson’s observations. Therefore this convention has implication for languages with co-articulated stops [kp], or glottalized consonants [p\(^\#\)], nasalized vowels [an], and so on.”

The outer square brackets around the palato-alveolar affricate [[t][ʃ]] seem to indicate that St. Clair regards it as an allophone of the segmental phoneme /tʃ/.

This convenient notational device of bracketing a segment as well as its subsegmental constituents facilitates the representation of paradigmatic and syntagmatic allophones. A paradigmatic allophone is defined here as one of a set of either segments or subsegments in a relation of mutual substitution; a syntagmatic allophone consists of a subsegment entering into sequence with other subsegments.
Taking lexical tone as an illustration, consider the plain voice High Level tone of the Mpi word si ‘four’ in Table A.1. To simplify matters, let us assume that this particular token is the prototypical plain voice allotone of the High Level toneme. The first tonal subsegment \( t_{1p} \) displays a rise 147-160 Hz, the second tonal subsegment \( t_{2p} \) a fall 172-158 Hz. Thus the plain voice High Level toneme is realized as the sequence of tonal subsegments \([t_{1p}][t_{2p}]\)\(_{HL}^{lev}\). As a result of their combination, the tonal subsegments \( t_{1p} \) and \( t_{2p} \) are syntagmatic allotones. Now let us turn to the tense voice High Level tone of the Mpi word si (name) in Table A.2. There is only one subsegment \( t_{2} \), a rise 158-167 Hz. Let us assume this token \([t_{1t}]\)\(_{HL}^{lev}\) is the prototypical tense voice allotone of the High Level toneme. Since the plain \([t_{1p}][t_{2p}]\)\(_{HL}^{lev}\) and the tense \([t_{1t}]\)\(_{HL}^{lev}\) allotones can be substituted for each other within the same segmental slot, they are paradigmatic allotones of the High Level toneme \(/[t_{1}]\)\(_{HL}^{lev}\).

Although further elaboration of the method is beyond the scope of this work, this highly simplified account shows that complex segments can be analyzed using established phonological principles. Remark also that the analysis adheres to the phonetic facts. In order to avoid coarticulatory effects, non-pitch speech parameters are most likely specified at the midpoint frame of each subsegment (as defined in Section 4.6), with the initial and final measures possibly playing a backup role. The fundamental frequency is sampled at the initial and final frames of each subsegment, if voiced. Thus, as far as the non-pitch speech parameters are concerned, the midpoint frame is the most representative of the subsegment as a whole, and accordingly, is the most probable bearer of the subsegmental phonological feature matrix.
This view is in harmony with proposals whereby each subsegment is assigned a column of phonological features, in effect expanding St. Clair’s bracketed phonetic symbol into a subsegmental phonological feature matrix (Hoard, 1967; Campbell, 1974; Sasse, 1976; Anderson, 1976, 1978a; Hoard, 1978; see also Ewen, 1982 from the perspective of dependency phonology). Hoard (1978), for example, defines a complex segment as “a segment with internal sequential structure and is representable as two or more columns of distinctive features.” In his general discussion of prenasalized and postnasalized stops, Anderson (1976: 335) points out the inherent difficulty in seeking to apply the features [+prenasal] or [+nasal] to the Kaingang medio-nasalized stops [(b)[m][b]] and [(d)[n][d]] (in St. Clair’s notation), which occur as allophones of prenasalized stops between oral vowels. Instead of positing the very doubtful feature [+postnasal] “to maintain the fiction that all feature specifications take the segment as their domain,” he finds it “far preferable to recognize smaller domains, and to treat complex nasal consonants as having internal structure analogous to the internal structure of complex tones.” Like Hoard, he later suggests that each subsegment be represented by a column of phonological features.

Two recent papers address the phonology of complex segments. Downing (2005) re-examines the familiar “one phoneme or two” question of nasals in homorganic NC sequences. Noting that cross-linguistically “there is often no difference in timing between between NC sequences analyzed as clusters and those analyzed as pre-nasalized segments, she states that “the decision about the segmental status of NC is necessarily motivated by phonological evidence.” In contrast to Fijian NC sequences that pattern phonetically and phonologically like unit segments (Maddieson and Ladefoged, 1993),
Downing prefers a cluster analysis of NC sequences in Bantu languages. Observe that the alternative phonological analyses can be easily accommodated by means of the notation /[N][C]/ for the one-segment sequence and /N//C/ for the two-segment sequence.

Within the framework of a study on Piro affricates, Lin (2005) summarizes the three hypotheses regarding the much debated phonological representation of affricates (some references omitted):

“The first treats affricates, at both phonological and phonetic levels, as contour segments that contain ordered stop and fricative units or ordered [–cont] and [+cont] specifications (Sagey, 1986). On the second view, the [–cont] and [+cont] components of affricates are phonologically unordered but become ordered phonetically (Hualde, 1988; van de Weijer, 1996). The third view is referred to the Stop Hypothesis: phonologically affricates are strident stops or stops contrasting with plain stops in place features but a fricative release or the [+cont] feature is added at the phonetic level (Jakobson et al., 1952; Steriade, 1993; Kehrein, 2002).

All three viewpoints set out by Lin are concerned with the phonological behavior of a unit segment such as /tʃ/, and not the two-segment sequence /tʃ/. The proposals differ, however, in the degree to which the ordered subsegments [t] and [ʃ] are visible to phonotactic rules. If the stop and fricative of the affricate are perceived in a distinguishable order according to the first pattern (Sagey), then the phonological visibility of both subsegments is expected to be high. On the opposite pole, if the affricate is considered an unanalyzable strident stop according to the third pattern (Jakobson et al.), then the subsegmental fricative should be largely invisible to the phonology. Piro is apparently an intermediate case since it displays phonological behavior similar to the first pattern under certain conditions and the third pattern under others. It seems reasonable to assume that affricates with a relatively brief fricative subsegment tend to follow the third pattern more than the first. If this suggestion is confirmed, then the “strident stop” of
Jakobson et al. (1952: 23-26) could be represented in a rather simple fashion by indicating an extra-short fricative in /\[t][ʃ]/ by means of the IPA breve diacritic “, for example (see Note 3 at the end of this chapter).

To conclude this section, Trubetzkoy’s decision to treat the aspirated and unaspirated stops in a fashion parallel to the voiceless and voiced stops is part of an enduring tradition in which all feature specifications are assumed to take the segment as their domain. Yet as Anderson (1976) noted in his discussion of Kaingang medio-nasalized stops, this assumption can lead to serious difficulties in phonological description when there are more than two subsegmental constituents. Sasse (1976: 146-147), for instance, gives a number of illustrations of phonemes with three subsegments although he considers four or more subsegmental constituents to be all but frequent, like the five subsegmental Korana Hottentot “lateral click-affricate || with an aspirated velar affricate accompaniment”: stop + || + k + x + h.\(^1\) Thus it is very likely that the subsegment is the appropriate domain of feature specification. From the point of view of speech processing, the subsegment is also the more probable unit of feature specification because the acoustic parameters are sampled within a smaller and hence more phonetically homogeneous domain.

7.1.3 Jakobson, Fant, and Halle (1952)

Of the twelve binary phonological oppositions put forth in the *Preliminaries to Speech Analysis* (Jakobson et. al., 1952: 40), two pertain directly to the glottal source: voiced/unvoiced and checked/unchecked. The voiced/unvoiced contrast is described as one opposing “buzz” (periodic vibrations of the vocal folds) to “hiss” (a noise source or
voicelessness in this work). The checked/unchecked contrast opposes sounds with an air stream checked by the closure of the glottis to sounds not generated in this way. The checked/unchecked contrast is illustrated by Circassian ejective and plosive stops: /p’a/ ‘place’ vs. /pa/ ‘be out of breath!’ Ejective fricatives are also mentioned.

Trubetzkoy’s contrast between aspirated and unaspirated sounds is not included among the twelve binary oppositions. However like the aspirated/unaspirated contrast, the proposed checked/unchecked opposition obscures the fact that an ejective stop may consist of two phonetically homogeneous subsegments (cf. the discussion in the preceding section). For example, Ladefoged and Maddieson (1996: 78-81) cite evidence that the Navaho velar ejective is produced with a long glottal closure after the supraglottal release: /[k’][?]/ using the notation adopted here. Observe its close formal resemblance with the aspirated voiceless velar plosive or the “voiceless velar plosive realized with a long VOT”: /[k][h]/.
<table>
<thead>
<tr>
<th>Voice Onset Time</th>
<th>Voicing leads</th>
<th>Voicing coincides substantially</th>
<th>Voicing lags moderately</th>
<th>Voicing lags considerably</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tense</td>
<td>No</td>
<td>Yes, if glottal constriction</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Voice</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Heightened subglottal pressure</td>
<td>Yes, if aspirated</td>
<td>Either</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Glottal Constriction</td>
<td>No</td>
<td>Yes, if HSGP; otherwise, optional</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

Table 7.1 The four VOT stop contrasts cross-classified by source features, following Table 8 in Chomsky and Halle (1968: 328); HSGP = heightened subglottal pressure.

7.1.4 Chomsky and Halle (1968)

By comparison to the more acoustically oriented approach of the earlier *Preliminaries to Speech Analysis* (Jakobson et. al., 1952), the binary feature system of *The Sound Pattern of English* (Chomsky and Halle, 1968) marks a return for the most part to the traditional motoric description of the positions and movements of the speech organs. However it is crucial for a system of phonological features to take into account both the motor and the acoustic aspects of speech communication. Reiterating the conclusions reached at the end of Chapter 2, not only does there appear to be a three-by-three motor hierarchy of phonation type (Section 2.4.2.3), but there is also evidence for an acoustic ordering of the nine phonation types along a linear scale of derived glottal bandwidth GBW (Section 2.4.2.4). Thus to explain patterns of phonological behavior, the
two sides of the Saussurean signifier, the motor and the acoustic, must be understood and formalized (cf. Flemming, 2002; see also van Reenen 1982, Chapters 1-2 for a thorough discussion of the motor and acoustic aspects of phonological features).

Chomsky and Halle (1968: 327) propose four source features:

1) Voicing
2) Tenseness
3) Glottal constriction
4) Subglottal pressure

They then distinguish four categories of stops, each associated with a different voice onset time (VOT): 1) voicing leads the release, 2) voicing coincides substantially with the release, 3) voicing lags moderately the release, 4) voicing lags considerably the release.

In Table 7.1 the four VOT stop contrasts are classified according to the source features. The three rightmost columns (voicing coincides substantially, voicing lags moderately, voicing lags considerably) are placed respectively in correspondence with the Korean fortis, lenis, and aspirated stops discussed earlier in Section 7.1.2. Although no examples of source feature matrices are provided, presumably an unaspirated bilabial stop [b] with voicing throughout the supraglottal closure is associated with the feature set of the first column [–tense, +voice, –heightened subglottal pressure, –glottal constriction], whereas an “aspirated p” or a “p realized with a long VOT” is assigned the feature set of fourth column [+tense, –voice, +heightened subglottal pressure, –glottal constriction].

Trubetzkoy’s aspirated/unaspirated contrast is conspicuous by its absence among the twelve binary features of Jakobson et al. (1952). Chomsky and Halle, on the other
hand, distinguish two types of aspirated voiceless stops, one with a short VOT (e.g. Korean lenis stop) and another with a long VOT (e.g. Korean aspirated stop). They indicate that the Korean fortis stop corresponds to the set of features in the second column (“voicing coincides substantially”), but it is not stated explicitly whether the feature set of the second column can or should be used concomitantly with ejective obstruents, that is, the “checked” sounds of Jakobson et al. (1952); the [ejection] or alternatively [glottal pressure] feature is treated separately on pages 323-324. Their [tense] feature is discussed at length. They mention a cineradiographic study by Perkell in which “there was a significant increase in pharynx width when nontense [d] was articulated but not when tense [t] was articulated.” However they do not provide any empirical data supporting the claim that stops with a short VOT (third column) are [–tense] while stops with a long VOT (fourth column) are [+tense]. There is a similar problem with the feature [heightened subglottal pressure]. In her *Survey of Phonological Features* (1988: 17), Keating gives the following brief assessment:

“As Sommerstein (1977) notes, the SPE proposal for voicing (the features [voice], [tense], [heightened subglottal pressure], and [glottal constriction]) was never widely accepted. The feature [glottal constriction] played a marginal, optional role for voicing; the feature [tense] described supralaryngeal adjustments for consonants that would suppress voicing, [heightened subglottal pressure] indicated extra energy for aspiration, and [voice] indicated a particular vocal cord configuration suitable for voicing, not vocal fold vibration per se. Perhaps their most important innovation in this respect was the decision to represent the glottal configuration but not the results of the configuration—i.e. not whether vocal fold vibration actually occurs. The various features interacted in somewhat complicated ways in determining vibration, aspiration, etc., and were therefore perhaps too hard to learn to use, rather than theoretically unacceptable; there was also little evidence presented in their support.”
In an attempt to provide a classification of all phonological oppositions involving the activity of the larynx, Halle and Stevens (1971: 201-202) propose four binary laryngeal features:

1) “Spread glottis. By rotation and displacement of the arytenoid cartilages, the vocal cords can be displaced outward relative to their positions for normal voicing, leaving a large glottal width. If the vocal-cord stiffness is sufficiently large, the combination of wide glottis and stiff glottal walls inhibits vocal-cord vibration. On the other hand, slackening of the glottal walls by reducing the

| Table 7.2 Classification of obstruents, glides, and vowels in terms of the proposed laryngeal features of Halle and Stevens (following their Table IX-5, 1971: 203); b₁ represents the “lax voiceless stop” of Danish, pᵲ is the “partially aspirated” lenis Korean stop. |
|---|---|---|---|---|---|---|---|---|---|
| obstruents | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| b₁ | b | p | pᵲ | bʰ | pʰ | ʰ | ?b | pᵲ |
| glides | w, y | | | ʰ | h, W, Y | ? | ? | ?w, ?y |
| vowels | V | Ṽ | V̂ | voiceless | breathy | creaky | glottalized |
| | | | | vowels | vowels | vowels | |
| spread | - | - | - | + | + | + | - | - | - |
| glottis | constricted | - | - | - | - | - | + | + | + |
| stιff vocal | cords | - | - | + | - | + | - | - | + |
| slack vocal | cords | - | + | - | - | + | - | + | - |

7.1.5 Halle and Stevens (1971)

In an attempt to provide a classification of all phonological oppositions involving the activity of the larynx, Halle and Stevens (1971: 201-202) propose four binary laryngeal features:

1) “Spread glottis. By rotation and displacement of the arytenoid cartilages, the vocal cords can be displaced outward relative to their positions for normal voicing, leaving a large glottal width. If the vocal-cord stiffness is sufficiently large, the combination of wide glottis and stiff glottal walls inhibits vocal-cord vibration. On the other hand, slackening of the glottal walls by reducing the
stiffness can lead to a condition in which vocal-cord vibration will occur, even with a relatively wide glottal opening.

2) **Constricted glottis.** Adduction of the arytenoid cartilages relative to the position for normal voicing (accomplished, perhaps, by fibers of the thyroarytenoid muscles, as well as by the lateral cricoartynoid muscles) can cause the vocal cords to be pressed together and the glottis to narrow or close. When the vocal-cord stiffness is large in this situation, vocal-cord vibration does not occur, and no air passes through the glottis. For a lower coupling stiffness, vocal-cord vibration can be initiated, probably with relatively narrow, peaked pulses.

3) **Stiff vocal cords.** Increasing the stiffness of the vocal cords makes the coupling between upper and lower edges of the vocal cords larger. Stiffening of the vocal cords affects glottal vibration, regardless of the size of the glottal aperture. When the vocal cords are in a configuration for normal voicing (neither spread nor constricted), the rate of vocal-cord vibrations increases with increasing stiffness. Increased stiffness of the vocal cords will inhibit vocal-cord vibration under the following circumstances: (a) when an obstruction in the vocal tract causes the intraoral pressure to build up and hence the pressure across the glottis to decrease; (b) when the glottis is spread to cause a wide aperture or when it is constricted. Thus an increased stiffness of the vocal cords tends to narrow the range of transglottal pressures and glottal apertures over which vocal-cord vibration occurs.

4) **Slack vocal cords.** The vocal cords can be made more slack by decreasing the coupling between upper and lower edges of the vocal cords. This is probably accomplished by a decrease in the tension of the vocal cords, as well as by a decreased stiffness of the walls of the glottis. Slackness of the vocal cords can allow glottal vibration to occur even with a spread or constricted glottis. When the vocal cords are slackened, there is a decrease in the frequency of glottal vibration.”

Although pitch and manner features are not explicitly distinguished by the authors, stiff and slack vocal folds make up an antagonistic pair of glottal frequency features whereas constricted and spread glottis form an antagonistic pair of glottal manner features. The combinations [+stiff, +slack] and [+spread, +constricted] are excluded a priori, yielding nine phonological categories out of a maximum of sixteen (2⁴), a fairly economical use of distinctive features compared to the earlier work of Chomsky and Halle (1968). The feature sets are claimed to be implemented quite differently according to whether they occur with obstruents, glides, or vowels. Table 7.2 illustrates their Table IX-5 on page 203.
Before turning to the comprehensive critique of this classification by Keating (1988), first of all observe that there is no straightforward way of indicating voice, as she points out below. This constitutes a significant failure of the feature system since the typological survey of voicing in Chapter 2 shows that the primary contrast in phonation type is between voiceless and voiced sounds (Section 2.4.2.2). The second major objection relates to treating a complex segment like the aspirated stop /p[h]/ as a unit for purposes of feature assignment. However the same objection can be leveled against all the feature proposals discussed in this chapter except the present one. In its favor, the feature system does capture the natural class of [h] and breathy vowels by means of [+spread]. For instance, Gujarati /bəhər/ (mēλə ‘outside’, Section 6.3.1) most likely becomes [bər] through an intermediate stage [bəhər], in which [a] undergoes the progressive assimilatory action of the [+spread] glottis of [h]. On the other hand, no evidence is provided indicating that voiceless vowels are produced by [–stiff vocal cords] but [h] is made with [+stiff vocal cords]. If this claim is accepted, voiced sounds in the environment of [h] would presumably have a higher pitch than those in a context of voiceless vowels.

Keating (1988: 17-22) offers a systematic review of the criticisms directed toward this feature system. To avoid a simple reformulation of her remarks, the major points are provided textually at the risk of some lengthiness:

“[These features] were meant to describe all aspects of laryngeal distinctions, including airstream mechanisms, phonation types, and fundamental frequency, as well as voicing and aspiration. The features attempted to relate all of these dimensions by representing them with shared dimensions. Sagey (1986) is a recent case in which the features are used in all of these ways. However, since the
original proposal, many criticisms of these features have been made (several of which are contained in chapters of the Fromkin, 1978 tone volume), both in terms of their physical phonetic accuracy, and their phonological usefulness. Some of these will be repeated here.

1. Voicing. First consider the Halle & Stevens features simply as a way of representing voicing distinctions. They do not offer a straightforward characterization in terms of whether or not the vocal folds are vibrating, even though they require a fairly precise physical description of the glottal state. Although Hayes (1984) suggests that some facts from Russian require this type of glottal representation, this move has not been generally endorsed. Such authors as Anderson (1974), Clements (1985), and even Stevens et al. (1986) include a separate [voice] feature, a cover feature to refer either to something like the presence or absence of vocal cord vibration, or to more abstract phonological categories.

Among phoneticians, the H&S features have not been widely adopted. Most importantly, claims made by H&S about vocal cord control for voicing have not been supported by experimental evidence. The H&S feature system assumes that voiced obstruents have slacked vocal cords to allow vibration, while voiceless obstruents have stiffened vocal cords to suppress it. However, Chen (1970) and Hirose and his colleagues (cited by Hombert et al. 1979) have provided evidence from laryngeal EMG that there is no difference in vocal cord stiffness for voiced vs. voiceless consonants.

The other H&S features [±spread glottis], [±constricted glottis] provide three degrees of glottal constriction (spread, neutral, constricted), as opposed to the many degrees used by Ladefoged or Catford. These two features have been widely adopted for the description of aspiration and glottalization, since there is a convenient feature value for each of these characteristics: [+] spread glottis for aspiration and [h], and [+constricted glottis] for glottal stops and glottalization. There is no doubt that the glottis is spread at consonant release for aspiration and constricted for glottalization, and thus these features values, taken apart from the system as a whole, are phonetically quite accurate. It is probably fair to say that many people use these two particular feature values as names for these characteristics without necessarily endorsing the whole system of four features. The features for vocal cord stiffness are, after all, the problematic ones.

2a. Tones and voicing. The basic idea is that stiff vocal cords raise F0 on a sonorant while slack vocal folds lower it; thus Low tone is represented by the combination [–stiff, +slack], Mid tone by [–stiff, –slack], and High tone by [+stiff, –slack]. Since slack vocal cords also facilitate voicing in obstruents, such voicing is associated with Low tone or lowered F0 in adjacent sonorant. Since stiff vocal cords also suppress voicing, voicelessness is associated with High tone or raised F0 in an adjacent sonorant. Since the vocal cords for voiced sonorants have neutral tension, they are associated with Mid tone, or no special pitch effects. Such correlations are indeed observed in languages, but not as uniformly as the hypothesis requires. One problem with this system is that in fact sonorants often affect F0 and tones, just as obstruents do. This is especially obvious when the sonorants come in voiced/voiceless pairs (Maddieson, 1984). Another point is
made by Traill et al. (ms) with respect to the ‘depressor’ consonants of Zulu. Here, where non-tonal effects of vowel F0 are phonologized, they are not correlated with phonetic voicing. Other problems along these lines are discussed by Anderson (1978b: 161-167) and Hombert (1978).

2b. Tones and other consonant dimensions. A further set of correlations with tone arises because vocal cord stiffness is also involved in describing phonation types and airstream mechanisms. Slack vocal folds not only imply Low tone and obstruent voicing; they are also used for breathy voice. Neutral vocal cords not only imply Mid tone and sonorant voicing; they also are used for implosives and two kinds of voiceless obstruents. Stiff Vocal cords not only imply High tone and voicelessness; they are also used for ejectives and voiceless glottalized obstruents. In qualitative phonetic terms, F0 differences of these kinds often are observed with these consonant types (though see Painter’s 1978 data for some counterexamples). One counterexample, even to these qualitative predictions, given by Hombert (1978) is that ejectives, unlike other glottalized consonants, are neutral with respect to tone. Other difficulties arise because the feature system does not confine itself to general qualitative predictions about ‘higher’ and ‘lower’ F0. Because it represents three different pitch levels, it makes much stronger, more specific claims that can run into empirical difficulty. For example, both plain voiced and breathy voiced consonants are predicted to lower tones and F0; they do, but in fact breathy voicing lowers F0 more than plain voicing.

2c. Tones and phonation type. Vowels themselves may vary in phonation type, and in the H&S system this should preclude orthogonal tone specifications. The H&S features predict the following correlations of phonation and tone: first, a plain vowel may have any of the three tones; second, breathy and creaky vowels, (as well as voiced consonants) go with Low tone; third, (plain) glides (and, vacuously, voiceless vowels) go with Mid tone; finally glottalized vowels go with High tone. However, in languages, various phonation types may occur with many different tones; for some examples of a single phonation type contrasting in tone, see Anderson (1978b), Ladefoged (1983) and Maddieson and Ladefoged (1985).”

At the beginning of the last section of her review, Keating (1988: 21-22) points out that the glottalic airstream mechanism is characterized as [constricted glottis] in the Halle and Stevens feature system. Thus the airstream mechanism is not directly encoded but is represented by the simultaneously occurring glottal stricture. There is reasonably good evidence in support of the premise that the acoustic consequences of a constricted glottis play an important part in keeping ejectives apart from plosives. As noted earlier, Chomsky and Halle (1968) propose the feature [ejection] or alternatively [glottal
In a similar vein, Ladefoged and Maddieson (1996: 372) put forward a Glottal Movement feature with two values: Raising and Lowering. Nevertheless, Kingston (1985, Chapter 2) shows that larynx raising alone appears not sufficient to produce an ejective; other articulatory maneuvers such as tongue root retraction and stiffened vocal tract walls may be required as well. Assuming larynx elevation to be the main motor correlate of ejectives, then the F1 of ejectives should be higher than the F1 of non-ejectives due to a shorter vocal tract and narrowed pharynx. Inspection of the Quechua supraglottal release data in Chapter 6 reveals that only the palato-alveolar ejective (t)’ has a higher midpoint F1 than the corresponding unaspirated and aspirated series (Table 6.4.4). The midpoint F1s of the uvular and velar ejectives do not conform to this hypothesis at all (Tables 6.4.1-6.4.2). By contrast, the glottal stricture measure PE1 distinguishes among the Quechua stop series more consistently. Hence it seems likely that the most important cue for ejectives is a constricted glottis although much more data is needed to confirm this conclusion.

As mentioned previously, Halle and Stevens consider the complex segment to be a unit for purposes of feature assignment. Yet the utterance-initial velar ejective /[k’][ʔ]/, for instance, consists of a minimum of two distinct subsegments with a constricted glottis, the first being the supraglottal release [k’], the second being the glottal closure [ʔ]. The duration of [ʔ] in the ejective /[k’][ʔ]/ may vary just as the duration of [h] does in the aspirated velar plosive /[k][h]/ (cf. Ladefoged and Maddieson, 1996: 79 for the short glottal closure of Hausa ejectives and the long glottal closure of Navaho ejectives).
Failure to recognize that ejective obstruents may be complex segments has led to some confusion in the analysis of their phonological behavior. Fallon (2002: 75-76) noted an asymmetry between [voice] and [constricted glottis] (see also Lombardi, 1994). He remarked that “the individual laryngeal feature [voice] can spread independently of [constricted glottis], as we saw in Waata Oromo, and in Kabardian and other NWC varieties,” but he did not find “any clear case of ejection spreading alone.” Moreover, he states his “inability to document the existence of such a change as /tk’/ → [t’k’] without also /tg/ → [dg].” Fallon overlooks the fact that ejectives are typically made up of two or more phonetically homogeneous subsegments. More realistic representations of his sound changes are: /[t]//[k’][?]/ → [[t’][?]][[k’][?]] and /[t]//[g]/ → [[d]][[g]], assuming the [?] subsegments to be of long duration. The relative rarity of the first assimilatory process may be due to the rather involved set of phonological operations visibly needed to implement it. 4

According to Keating, one of the principal weaknesses of the Halle and Stevens feature system is the positing of fixed associations between a tone and a phonation type. 5 As she indicates, a good deal of counterevidence has accumulated against this claim. Nonetheless, creaky voice (low pitch register) is quite often associated with considerable modulation noise or harshness (cf. the final log pACC of the Mpi tense voice low tone in Table A.10; see also Michel, 1964). During creaky voice not only are the true vocal folds adducted, but the ventricular and perhaps the aryepiglottic folds are brought together as well (Section 2.2.2). The ventricular and aryepiglottic folds are most likely recruited in order to increase the effective mass of the vocal folds, thereby lowering the fundamental frequency. Yet the resulting spatial nonuniformity of the vocal folds also creates a
nonlinear vibrating system, which in turn can give rise to a significant amount of modulation noise (Section 3.2.1). Therefore, the frequent association of harshness with creaky voice probably results from the strong tendency of the glottal manner mechanism (constricted vocal folds) to act in synergy with the main glottal frequency mechanism (slack vocal folds). A synergistic relationship between two relatively independent speech parameters is far from being a rare phenomenon. Stevens and his colleagues have examined how certain featural properties can enhance or strengthen others among a number of different phonetic dimensions (Stevens et al., 1986; Stevens and Keyser 1989; Keyser and Stevens, 2001).

The relative independence of tone and phonation type is but one example showing how crucial it is not to build fixed associations into the feature system. On the other hand, there needs to be a way of stating possibly phonetically motivated feature associations when they do occur, whether in the individual languages or as a typological tendency. The usual formalism for representing co-occurrences between features is a “redundancy” rule in the form of a conditional if-then statement $a \rightarrow b$, where $a$ is the antecedent (the “if” part of the statement), $b$ the consequent (the “then” part), and the arrow $\rightarrow$ the conditional operator (Stanley, 1967). A more explicit if-then statement is used by Peng
as well as Archangeli and Pulleyblank (1994). Thus in the Halle and Stevens feature system, the conditional association of harshness with creaky voice could be expressed either as: [+slack vocal folds] → [+constricted vocal folds] or as: if [+slack vocal folds], then [+constricted vocal folds].

Not only does tone interact on occasion with the different phonation types, but vowel height appears to do so as well. In an often cited review of Mon-Khmer register, Gregerson (1976) presents the two registers and their related characteristics (Table 7.3).

<table>
<thead>
<tr>
<th>Initial (written) Voice Vowel</th>
<th>Consonant Quality</th>
<th>Vowel Quality</th>
<th>Pitch</th>
</tr>
</thead>
<tbody>
<tr>
<td>First Register</td>
<td>(original) surds</td>
<td>normal head tense</td>
<td>more open, onglided</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Second Register</td>
<td>(original) sonants</td>
<td>deep breathy sepulchral chest relaxed</td>
<td>close, centering diphthongs</td>
</tr>
</tbody>
</table>

**Table 7.3** The two Mon-Khmer registers and their related characteristics according to Gregerson (1976: 323).

(1992) as well as Archangeli and Pulleyblank (1994). Thus in the Halle and Stevens feature system, the conditional association of harshness with creaky voice could be expressed either as: [+slack vocal folds] → [+constricted vocal folds] or as: if [+slack vocal folds], then [+constricted vocal folds].

Not only does tone interact on occasion with the different phonation types, but vowel height appears to do so as well. In an often cited review of Mon-Khmer register, Gregerson (1976) presents the two registers and their related characteristics (Table 7.3). In the first register a tense voice quality corresponds to a lowered vowel (higher F1) while in the second a breathy or relaxed voice quality is associated with a raised vowel (lower F1) and perhaps depressed larynx (see Wayland and Jongman, 2003 for an acoustic study of Khmer clear and breathy vowels). Furthermore, he observed that certain West African languages exhibit parallels with the Mon-Khmer register system. The lowered vowels of Twi and Fante are respectively “strangled” and “creaky” whereas the raised vowels of the two languages are respectively “hollow” and “breathy”. The contrasting vowel heights of both language groups are largely attributed to a difference in
the volume of the pharyngeal cavity, lowered vowels being characterized by a constricted pharynx, raised vowels by an expanded pharynx.

Following up on Gregerson’s observations, Denning (1989) selected a sample of fifty languages from about seventeen genetic language groups. The universal tendencies he found are illustrated in Table 7.4. Summarizing, he states that (1989: 67):

“Despite exceptions, the pattern stated in the universal is supported by the bulk of the data and represented over a much more geographically and genetically diverse range than its converse. Of all languages which have been described as having a regular correlation of relative vowel height and voicing or phonation type, those which show a positive correlation between greater height and the concomitants of laryngeal laxness (i.e. breathiness, voicing or lower pitch) occur with roughly four times the frequency of those which do not.”

Note, however, that the hypothetical universal associating high and low vowels with respectively low and high fundamental frequencies appears rather doubtful. A survey of 31 languages has shown that high vowels consistently have a higher fundamental
frequency than low vowels (Whalen and Levitt, 1995; but see Esling, 1999b on the low pitch of “lowered larynx voice” and the remarks of Denning, 1989: 93-94 addressing this problematic issue).

After identifying the various biomechanical, neuromuscular, aerodynamic, and acoustic factors involved, Denning (1989, Chapter 4) concludes that they all contribute in varying degrees to the observed correlation between vowel height and phonation type. The most attractive line of inquiry seems to be his neuromuscular hypothesis in which a constricted pharynx is accompanied by a constricted glottis (Denning, 1989: 86-87). Accordingly, a lowered vowel with a small-volume pharyngeal cavity (higher F1) may give rise to tense or harsh voice. Laufer and Condax (1979) conducted a fiberscopic study of Hebrew pharyngeals and found that (cf. also Ghazeli’s remarks in Section 2.4.2.3):

“Anatomical considerations may explain involuntary glottal stop and creaky voice connected with the production of pharyngeals. The muscles that pull the epiglottis down and back (fold it at approximately the middle of the cartilage) are the thyroepiglottic muscle and the aryepiglottic muscle. The thyroepiglotticus pulls the edge of the epiglottis down and back and possibly somewhat laterally, while the aryepiglotticus pulls it down and back and medially; this same muscle continues onto the larynx as the oblique interarytenoid. Contraction of the aryepiglotticus not only folds the epiglottis down, but the combined action of these two muscles can act to pull the epiglottis and the arytenoids toward each other. …the force of the muscular contraction is fully on the arytenoids, tending to pull them more tightly together than they would be if the epiglottis were not involved. This leads to creak, and also to what we referred to earlier as an involuntary glottal stop. The action of forming an epiglottal fricative or stop uses to some extent the same gestures as does swallowing.”

Painter (1986) notes the occurrence of a constricted glottis during the nonspeech gestures of effort closure and swallowing:

“In effort closure, the vocal folds are adducted, the false folds are also brought together and the entry to the larynx is obliterated by approximating the cuneiform cartilages and aryepiglottic folds and by retracting the epiglottis.
…In swallowing airway protection is needed and we note in particular that the epiglottis is drawn backwards over an already closed airway.”

He then surveys the world’s languages and points out a number of pharyngeal or pharyngealized speech sounds (small-volume pharyngeal cavity) that use similar effort and swallowing gestures.

In a parallel, though less well attested fashion, a raised vowel (lower F1) with a large-volume pharyngeal cavity may lead to the generation of lax or breathy voice, as “lowered larynx voice” is associated with an open laryngeal vestibule and breathiness (Esling, 1999b). Citing the work of Shin et al. (1981), Trigo (1991: 133) states that “opening the glottis involves the intrinsic laryngeal musculature and at least one of the infrahyoid muscles that lower the larynx (sternothyroid).” However, she also mentions that both the glottalized and breathy vowels of !Xôô can be either pharyngealized or non-pharyngealized (Traill, 1985: 68). Hence glottal stricture and pharyngeal cavity volume can be controlled independently of each other despite their frequent cross-language association.
An example of a relatively recent glottal manner feature system is the one proposed by Kenstowicz (1994: 40-41) in Table 7.5. Note the evident continuity with the feature system of Halle and Stevens (1971). The feature [voice] is an important innovation because it fills a serious gap in the Halle and Stevens system, as discussed earlier. Iverson (1983) also puts forward a feature system similar to the one in Table 7.5 although it differs in some details and lacks the [voice] feature. Complex segments continue to be treated as a single unit rather than as two or more subsegments, each bearing its own feature specification, such as the ejective /pʰ/ /[p’][ʔ]/ when [ʔ] is of relatively long duration, the aspirated voiceless stop /pʰh/ /[p][h]/, and possibly the aspirated voiced stop /bʰ/ /[b][fi]/, where [fi] indicates the breathy voiced sound.\(^7\)

Glottal and supraglottal noncontinuants are represented by a maximum of three phases: attack, closure, and release (see Section 2.4.1). Accordingly, the generic stop /P/ may be formed by three successive phonetically homogeneous subsegments /[P]_a[P]_c[P]_r/.

<table>
<thead>
<tr>
<th></th>
<th>Spread</th>
<th>Constricted</th>
<th>Stiff</th>
<th>Slack</th>
<th>Voiced</th>
</tr>
</thead>
<tbody>
<tr>
<td>(p^h) (Hindi)</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>(b^h) (Hindi)</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>(p) (English)</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>(b) (French)</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>(p^*) (Korean)</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>(\delta)</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

**Table 7.5** The laryngeal features proposed by Kenstowicz (1994: 40-41). The fortis Korean stop is indicated by \(p^*\).
where [P]_a is the attack transient, [P]_c the closure, and [P]_r the release transient or burst. Hence utterance-initial /p/ and most likely /p*/ (the Korean fortis stop) consist of just one subsegment: the release bursts /[/p]_r/ and /[/p*]_r/. On the other hand, utterance-initial voiced stops /b/ and /b/ may constitute the complex segments /[/b]_c[/b]_r/ and /[/b]_c[/b]_r/ on the condition that the release transients [b]_r and [b]_r are acoustically distinct.

Kenstowicz’s feature framework represents a major improvement over the Halle and Stevens system because fixed associations involving tone and voice are no longer explicitly posited, the main thrust of Keating’s 1988 critique. Yet no evidence is offered as to why the implosive /b/ is considered [+slack], for example, while Halle and Stevens as well as Iverson (1983) prefer [−slack]. Following Catford (1977: 108), one often has the impression that earlier feature systems are subject to “the procrustean forcing of items into particular categories, whether this categorization corresponds to reality or not.” To avoid excessive abstraction, the present work has sought to ground its premises concerning the different phonation types on auditorily motivated parameters obtainable from the speech wave. Furthermore, a set of criteria for determining the boundaries of phonetically homogeneous subsegments is proposed in Chapter 4 (Section 4.6). This approach demands, however, much more attention to measurable phonetic detail in the language under investigation than is customary on the part of many phonologists.

Another relatively recent glottal manner feature system is the one developed by Lombardi (1994). After reviewing obstruent voicing contrasts in a number of languages, she concludes that a feature system must account for the following series (1994: 18-19):

a. voiced glottalized (usually implosive)
In light of the fact that the phonation type of coda obstruents is often neutralized in favor of voicelessness, Lombardi hypothesizes that that glottal manner features are privative and have no negatively specified values. To make the necessary distinctions in the obstruent series, three privative features are consequently adopted: [voice], [glottalization], [aspiration]. The obstruent series are cross-classified by the three features in Table 7.6.

Observe the close similarity of her model to the system proposed by Kenstowicz in Table 7.5. The use of the pitch features [slack] and [stiff] allows him to make finer distinctions within the voiced series (English b vs. French b) and the constricted series (p̃ vs. Korean fortis p*), but if the pitch distinctions are eliminated and the minus signs removed, his system becomes identical to Lombardi’s.
The choice of the term “aspiration” over “spread” gives Lombardi’s feature system an added clarity since the word refers unequivocally to the second subsegment of /p[/ and /b[/ (see Lombardi, 1994: 7-10 for evidence of voiceless and voiced aspirated sounds pattern together phonetically and phonologically). The term “spread”, on the other hand, is somewhat ambiguous in the feature systems of Halle and Stevens (1971), Iverson (1983), and Kenstowicz (1994) because the term may indicate the state of the glottis during either the plosive subsegments [p], [b] or the aspirated subsegments [h], [fi]. The term “constricted” applied to an ejective /p’,?/ suffers from the same problem in the above feature systems because it can indicate the glottal state during either the supraglottal release [p’], subsegment or the following [?] subsegment.

Several similar glottal manner systems with two or three features have been introduced in the last decade or so. A brief comparison is given in Table 7.7 (see Honeybone, 2005 for further discussion of this table and references therein). The principal innovation by Lombardi (1994) as well as Iverson and Salmons (1995) has been the exclusive use of privative as opposed to binary features. However, two major difficulties arise in the application of a privative feature system. The first concerns the

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>[aspiration]</td>
<td>H</td>
<td>[spread glottis]</td>
<td>[+tense]</td>
<td></td>
</tr>
<tr>
<td>[voice]</td>
<td>L</td>
<td>[voice]</td>
<td>[+voice]</td>
<td></td>
</tr>
<tr>
<td>[glottalization]</td>
<td></td>
<td>[constricted glottis]</td>
<td>[+checked]</td>
<td></td>
</tr>
</tbody>
</table>

Table 7.7 A comparison of recent glottal manner systems after Honeybone (2005).
phonetic interpretation of privative features, the second relates to the phonological validity of universal privative features.

To provide a framework for the phonetic interpretation of their binary feature system, Chomsky and Halle (1968: 166) define a phonetic feature specification as being made up of 1) a phonetic scale and 2) “an integer indicating the position of the phonetic segment in question along this scale.” In a footnote on the same page, they continue:

“Often we restrict ourselves to two positions along a phonetic scale, in which case we may use the symbols + and – instead of integers to indicate phonetic values. We emphasize that the value in a phonetic specification is not an absolute physical property but is relative to the context of phonetic segments.”

The dependence of phonetic feature specifications on the environment has already been discussed in Chapter 4 (Section 4.3), in which it is pointed out that the values of the modulation noise parameter (log pACC) can be compared with each other only on the condition that the corresponding fundamental frequencies are roughly the same. Fant (1973: 210) provides further observations on contextual dependence:

“The phonetic value of a distinctive feature can be regarded as a vector in a multidimensional signal space. The variability due to context shall be expressible by rules which define how the feature vector is changed when the conditioning elements are varied. A minimum requirement of phonetic reality of a feature would be that any two sounds in any context differing by one and the same feature only shall display a difference vector of the same sign along a common phonetic dimension.”

Lending more substance to Chomsky and Halle’s brief treatment of phonetic feature specification, Anderson (1974: 9-10) refines their concept of the phonological division of a phonetic scale and offers an example to illustrate it (one reference omitted):

“Thus we will assume that features have binary values, but that underlying these values is a specification of each parameter in more detailed terms, which is given on a continuous real-valued scale which we will arbitrarily assume to vary between 0 and 1. This scale is interpreted in such a way that values below some given point are interpreted as –F, while higher values are interpreted as +F. The
dividing point for a given feature may well vary from language to language, and language-particular rules may well affect the detail values as well as the binary values of features. We will see, subsequently, some interesting differences between the rules specifying binary values and those interpreting or specifying these as real-number values, which we will refer to as detail rules. An example of the sort of thing referred to is the following. In the Breton dialect of Plougrescant, vowels can be distinctively either nasal or nonnasal. Thus, phonetic representations will specify some vowels as [+nasal], and others as [–nasal]. But among the oral vowels, some are more nasal than others: [–nasal] vowels immediately preceding nasal consonant are slightly nasalized, though not as much as distinctively nasal vowels. We thus have at least the following three values (where the precise values, on a hypothetical scale, are irrelevant; only the relative values matter): [0nasal] (for oral vowels next to oral consonants); [.3nasal] (say) for oral vowels next to a nasal consonant; and [.7nasal] for distinctively nasal vowels. We can then say that any value over [.5nasal] counts as [+nasal], while values below [.5nasal] count as [–nasal]. We then have three values (at least) phonetically, but only two values in the slightly more abstract binary representation. Furthermore, some instances of the property in question (lowered velum) are treated as ‘presence of F’ (i.e., [+F]), while other instances of the property if low enough in value, are treated as ‘absence of F’ (i.e., as [–F]). We presume the reality of both sorts of phonetic representation in linguistic description.”

At the conclusion of his paper on the glottal manner of fricatives, Vaux (1998) adopts a view of the phonological division of the phonetic scale very similar to Anderson’s:

“The important point to bear in mind here is that a phonological specification [–X], where X is any feature, does not entail that the component of the vocal tract activated by [X] is completely inert at the phonetic level. The English phoneme [ʃ], for example, is specified in the phonology as [–round], yet it is often implemented with some rounding of the lips. Similarly, vowels that are phonologically [–nasal] are often implemented with some airflow through the nasal passage by speakers of American English. The fact that in each of these cases a particular articulator is phonetically active to a certain degree does not entail that this activity is encoded in the phonological representation; for example, we do not want to say that English [ʃ] is [+round] in the phonology, nor do we want to say that American vowels are phonologically [+nasal]. Rather, we should acknowledge that phonological features bisect a continuum of phonetic activity. In this view, [+nasal] for example represents lowering of the velum beyond a certain critical zone; [–nasal] represents any lesser degree of velar lowering. Similarly, [±spread glottis] should be defined in terms of a line drawn somewhere in the range of possible degrees of spreading of the vocal folds [Figure 7.1.1],
rather than in the all-or-nothing terms of spreading versus no spreading [Figure 7.1.2].”

**Figure 7.1.1** Binary division of the \ [+spread glottis] \ phonetic scale, adapted from Vaux (1998).

**Figure 7.1.2** All-or-nothing (privative) division of the \ [+spread glottis] \ phonetic scale, adapted from Vaux (1998).
There is overwhelming evidence from the present study and other reports that the continuous real-valued phonetic scale is divided into two or more phonological categories (albeit often with fuzzy boundaries), as presented in Figure 7.1.1. As a result, the privative or all-or-nothing partition of the phonetic scale in Figure 7.1.2 is clearly incorrect from the perspective of speech processing.

For instance, the log pACC has been used successfully as the main indicator of speech periodicity in this work (see Sections 4.2.3 and Chapter 5 for details). If its value falls below the conditional threshold, then the speech waveform is judged [–voice], if not then [+voice]. Once the speech waveform is deemed [+voice], however, a low value of the log pACC serves as a cue for harshness (significant modulation noise or period-to-period irregularity). Because Jalapa Mazatec distinguishes between lexically contrastive voicelessness, harsh voice, and pure (nonbreathy) voice, the continuous phonetic scale of log pACC values is subject to a three-way phonological division (Section 6.3.2). In Mpi the phonetic scale of spectral tilt is bipartitioned into phonemic tense and plain voice (Section 6.2), and other instances of the phonological division of the phonetic scale could be drawn from Chapter 6 as well. Repp’s review article (1984) on categorical perception provides a number of examples whereby a continuous speech parameter is bisected or trisected into regions of enhanced discriminability.

To address the question of the phonological validity of universal privative features, Baltaxe (1978: 57-58) points out that Trubetzkoy considered his three categories of oppositions: privative, gradual, and equipollent to be interconvertible in certain cases:

“For example, the contrast between voiced and voiceless stops could be considered equipollent, that is, as having two contrastive features: the contrastive feature for the /d/b/g/ would be the vibration of the vocal cords, which is absent in the voiceless stops. The contrast feature for /p/t/k/ would be the tensing of the
buccal muscles, which is absent in voiced stops. The opposition between voiced and voiceless stops can also be considered privative. It becomes privative only when one of the features, vibration of the vocal cords or tensing of the buccal musculature, is abstracted and considered distinctive. Trubetzkoy outlined guidelines to convert this type of opposition: ‘…to interpret (these) relations as privative, attention must be focused on one property, and the lesser degree thereof must be equated with zero’…’

Baltaxe further remarks (1979: 59):

“Thus, after Trubetzkoy has abstracted the three categories he did not consider assignment of a contrast to any one of them as absolute. It was possible to change within these categories. However, a final choice was not arbitrarily made in the classification of phonological oppositions. The choice depended on the phonological system of a given language.”

In view of the continuous nature of the phonetic scale, very compelling evidence should be offered to support a privative feature analysis.

To determine whether privative [voice] is a language universal, Wetzels and Mascaró (2001) examined voice assimilation processes in a relevant language sample. If [–voice] is always phonologically absent and thus never plays a role in voice assimilation, then the hypothesis of universal privative [voice] holds. If, on the other hand, [–voice] is phonologically active and hence spreads to adjacent segments in a manner parallel to [+voice], then there is no reason to assume universal privative [voice]. Since several languages in the sample are characterized by phonologically active [–voice], it is clear that privative [voice] does not constitute a language universal. Languages with phonologically active [–voice], but with either inert or optionally active [+voice], include Yorkshire English, Parisian French, Ya:thê, as well as Makkan Arabic (cf. also Kim 2002). In sum, there appears to be no basis for universal privative [voice] since both [+voice] and [–voice] can be phonologically active in voice assimilation.
7.2 The present system of glottal manner features

7.2.1 The acoustic-motor classification of the nine phonation types

Before discussing the proposed system of glottal manner features, it is useful to recall the acoustic-motor classification of the nine phonation types presented under Section 2.4 of Chapter 2. In comparison with Table 2.2, note that the first formant quality factor (Q1) in Table 7.8 has been replaced by the first formant peak energy factor (PE1) defined in Chapter 4, tested in Chapter 5, and used successfully to distinguish among the three kinds of glottal noise in Chapter 6.

On the primary motor dimension of position I. (extreme ~ central-extreme ~ central),

A) the extreme position of the vocal folds is cued by glottal noise (aperiodic waveform, see Section 4.2.3),

B) the central-extreme position by noisy voice (periodic waveform; significant modulation noise, significant aspiration noise, or both),
the central position by pure voice (periodic waveform; neither significant modulation noise nor significant aspiration noise).

On the secondary motor dimension of stricture II. (constricted ~ constricted-spread ~ spread),

A) within glottal noise, 1) the constricted vocal folds are cued by a high PE1, 2) the constricted-spread folds by an intermediate PE1, and 3) the spread vocal folds by a low PE1;

B) within noisy voice, 4) the constricted vocal folds are cued by a low log pACC, 5) the constricted-spread folds by both a low log pACC and a low HNR, and 6) the spread vocal folds by a low HNR;

C) within pure voice, 7) the constricted vocal folds are cued by a flat spectral tilt, 8) the constricted-spread folds by an intermediate spectral tilt, and 9) the spread vocal folds by a steep spectral tilt.

<table>
<thead>
<tr>
<th>GBW</th>
<th>Phonation Type</th>
<th>Glottal Periodicity</th>
<th>Vocal Fold Position</th>
<th>Glottal Stricture Bandwidth</th>
<th>Vocal Fold Stricture</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Glottal stop</td>
<td>glottal noise</td>
<td>extreme</td>
<td>large</td>
<td>constricted</td>
</tr>
<tr>
<td>2</td>
<td>Whisper</td>
<td></td>
<td></td>
<td>medium</td>
<td>constricted-spread</td>
</tr>
<tr>
<td>3</td>
<td>Breath</td>
<td></td>
<td></td>
<td>small</td>
<td>spread</td>
</tr>
<tr>
<td>4</td>
<td>Harsh voice</td>
<td>noisy voice</td>
<td>central-extreme</td>
<td>large</td>
<td>constricted</td>
</tr>
<tr>
<td>5</td>
<td>Harsh whisphery voice</td>
<td></td>
<td></td>
<td>medium</td>
<td>constricted-spread</td>
</tr>
<tr>
<td>6</td>
<td>Breathy voice</td>
<td></td>
<td></td>
<td>small</td>
<td>spread</td>
</tr>
<tr>
<td>7</td>
<td>Tense voice</td>
<td>pure voice</td>
<td>central</td>
<td>large</td>
<td>constricted</td>
</tr>
<tr>
<td>8</td>
<td>Plain voice</td>
<td></td>
<td></td>
<td>medium</td>
<td>constricted-spread</td>
</tr>
<tr>
<td>9</td>
<td>Lax voice</td>
<td></td>
<td></td>
<td>small</td>
<td>spread</td>
</tr>
</tbody>
</table>

Table 7.9 The nine phonation types arranged 1 through 9 according to an increasingly narrower broad-spectrum glottal bandwidth GBW.
In addition to the three-by-three motor hierarchy of phonation type determined by the primary motor dimension of position (extreme ~ central-extreme ~ central) and the secondary motor dimension of stricture (constricted ~ constricted-spread ~ spread), typological evidence was adduced in Chapter 2 (Section 2.4.2.4) for an ordering of the nine phonation types along an linear acoustic scale of broad-spectrum glottal bandwidth that progressively narrows from GBW(1) glottal stop to GBW(9) lax voice, as shown in Table 7.9.

Remark that the fundamental frequency is well defined solely when the temporal periodicity of voice exists, or alternatively, the spectral narrowness of the glottal bandwidth GBW is equal to or greater than 4. Hence the pitch feature system is operant only under the glottal manner of voice.

7.2.2 The proposed system of glottal manner features

7.2.2.1 Equipollent motor features and voice assimilation

The proposed system of glottal manner features is illustrated in Table 7.10. Although [voice] and [noise] are acoustical terms, they are in fact being used as substitutes for the motor position features [central] and [extreme], respectively. These substitutions can be justified as follows:

1) [voice] and [noise] are the acoustic consequences associated with [central] and [extreme], as indicated in Tables 7.8-7.9.

2) The acoustic correlates of [voice] and [noise] can be directly estimated from the speech signal, whereas those of [central] and [extreme] can not.
3) [voice] is a well-known and long-established term.

As a consequence, the features of the primary motor dimension of position consist of [voice] and [noise] (or more exactly [central] and [extreme]). The features of the secondary motor dimension of stricture are simply the familiar [constricted] and [spread].

Taking into account Gallistel’s work (1980) on the organization of motor patterns, Halle (1983: 99-100) puts forward the analysis that phonological features are implemented through the coordinated action of opposing muscle groups:

“In characterizing movements of structures connected by joints, Gallistel observes that ‘a stimulus that excites a muscle on one side of a joint invariably inhibits excitation of the antagonistic muscle on the other side of the joint, and vice versa.’ While the active articulators are not connected to other parts of the vocal tract by joints, their movements are in many cases controlled by paired sets of agonistic and antagonistic muscles. Thus, for example, the lowering and raising of the velum under the control of the distinctive feature nasal [nasal] is implemented by the palatopharyngeus and palatoglossus, which together lower the velum, and the tensor veli palatini and levator veli palatini, which raise it. When the tensor and levator are excited and the palatoglossus and palatopharyngeus are inhibited, the velum is raised and no air can flow through the nasal cavities. When the former two muscles are inhibited and the latter two excited, the velum is lowered and air flows freely through the nasal cavities producing a specific acoustic effect which is referred to as nasalization.
The feature [coronal] is apparently controlled in much the same fashion as nasality [i.e., by a pair of agonist-antagonist muscles]. To produce a [+coronal] sound the tongue blade must be raised; to produce a [–coronal] the tongue blade must be lowered. Blade raising is implemented by contracting the superior longitudinal muscles of the tongue and relaxing the inferior longitudinal muscles; whereas blade lowering is produced by relaxing the superior pair of muscles and contracting the inferior pair.”

Summarizing, the positively specified feature [+F] excites the agonists (and inhibits the antagonists), whereas the negatively specified feature [–F] inhibits the agonists (and excites the antagonists). Halle (1983: 101-102) then considers the groups of muscles controlling vowel height and backness. His discussion of vowel height is of particular interest because the features [high] and [low] are widely accepted to be an equipollent pair. The feature specifications of the three vowel heights as well as the postulated excitatory (+) or inhibitory (–) patterns of the posterior genioglossus (PGG) and the hyoglossus (HG), following Halle (1983).

<table>
<thead>
<tr>
<th>Feature</th>
<th>High</th>
<th>Low</th>
<th>PGG</th>
<th>HG</th>
</tr>
</thead>
<tbody>
<tr>
<td>High vowel</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>Mid vowel</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Low vowel</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>+</td>
</tr>
</tbody>
</table>

Table 7.11 The feature specifications of the three vowel heights and the presumed excitatory (+) or inhibitory (–) patterns of the posterior genioglossus (PGG) and the hyoglossus (HG), following Halle (1983).

Remark the parallels between the feature specifications, whether positive or negative, and the presumed excitation patterns of the posterior genioglossus and the hyoglossus. Moreover, Halle (1983: 103) states that “…the posterior genioglossus is excited for [+high] sounds and inhibited for [–high] sounds, where the hyoglossus is excited for [+low] sounds and inhibited for [–low] sounds.”
Steriade (1995: 150) offers the proposal that equipollent features “correspond in fact to two distinct privative features defined as two opposing gestures on the same or related articulatory dimensions.” As only the agonist-excitatatory, positively specified +F value of each member of the equipollent pair $[F_1, F_2]$ is allowed under this hypothesis, the proposal effectively denies the phonological relevance of the agonist-inhibitory, negatively specified –F value. Yet the negatively specified features, [–high], [–low], and [–high, –low], can all be used to characterize natural classes. Kenstowicz and Kisseberth (1979: 246) provide some illustrations:

1) “[–high] vowels: e.g., in Russian the nonhigh vowels $e$, $o$, and $a$ undergo vowel reduction in unstressed position, while the high vowels $i$ and $u$ do not.

2) [–low] vowels: e.g., in Chamarro unstressed high and mid vowels reduce to $I$ (if front) or to $U$ (if back), while unstressed low $\ddot{a}$ and $a$ appear as $\varnothing$.

3) [–high, –low] vowels: e.g. in Lamba $i$ is lowered to $e$ after the mid vowels $e$ and $o$, but not after the high vowels $i$ and $u$, nor the low vowel $a$.”

Further examples are not difficult to find,

4) [–low]: “the most prominent environment that could induce palatalization in a consonant is a following front vowel, (especially the high- and mid-front unrounded vowels $i$ and $e$), and a following palatal semivowel (yod)” (Bhat, 1978: 60).

5) [–high]: “rounding harmony is favored when the trigger is non-high” (Kaun, 1995: 64).

On the basis of this evidence, it seems reasonable to conclude that each member of the equipollent pair of features $[F_1, F_2]$ has an agonist-excitatatory, positively specified +F value as well as an agonist-inhibitory, negatively specified –F value.

Although Steriade’s proposal to replace equipollent features by two privative ones is not borne out by the data, her proposal has the merit of calling attention to the fact that it is the agonist-excitatatory, positively specified feature +F that most often appears to be phonologically active or the trigger in assimilation phenomena, whichever feature +F₁ or
+F₂ of the equipollent pair [F₁, F₂] takes on this role in a given language (see Rice 1999a, 1999b for examples on how the choice of the trigger +F₁ or +F₂ can vary from language to language). Nevertheless, Steriade does point out that negatively specified [–high] can also act as a trigger, although apparently less often (1995: 156, references omitted):

“We must return now to the pattern of Bantu vocalic assimilation discussed earlier. Recall that mid vowels propagate [–high] onto suffixal high vowels, whereas the phonetically non-high a does not. The facts are repeated below:

<table>
<thead>
<tr>
<th>Past</th>
<th>Neuter</th>
<th>Applied</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>tul-a</td>
<td>tul-ika</td>
<td>tul-ila</td>
<td>“dig”</td>
</tr>
<tr>
<td>fis-a</td>
<td>fis-ika</td>
<td>fiš-ila</td>
<td>“hide”</td>
</tr>
<tr>
<td>kos-a</td>
<td>kos-eka</td>
<td>kos-ela</td>
<td>“be strong”</td>
</tr>
<tr>
<td>sek-a</td>
<td>sek-eka</td>
<td>sek-ela</td>
<td>“laugh at”</td>
</tr>
<tr>
<td>pat-a</td>
<td>pat-ika</td>
<td>pat-ila</td>
<td>“scold”</td>
</tr>
</tbody>
</table>

[There is the issue of] the binarity of [high], whose plus-value is the only one to spread in Romance. Only [+high] represents a significant deviation from the neutral position, hence we would expect only [+high] to be active. The facts of Bantu are problematic in that [–high] is clearly active here.”

An immediate problem that confronts any attempt to generalize equipollent features is that phonologically active [nasal] does not seem to have a phonologically active counterpart [oral]. Similarly, phonologically active [round] does not appear to have an active counterpart [drawn]. Steriade (1995: 149) mentions that “there exist processes possessing the appearance of local [–nasal] assimilation. Local postoralization (ma → mⁿa) and preoralization (am → aʰm) have been discussed in terms of spreading orality.” But apart from these isolated exceptions, she finds that “there is virtually no evidence left suggesting that orality is represented phonologically, in any language.”

It is questionable, however, whether motor assimilation should be the sole criterion on which to base the existence of a phonological feature. From the perspective of phonetics, the feature pair [nasal, oral] is just as equipollent as the height pair [high,
low]—if not more so—since the lowering and the raising of the velum is clearly
governed by paired sets of agonistic and antagonistic muscles, as illustrated by Halle
above. In addition, Merrifield (1963) found three degrees of lexically contrastive vowel
nasalization in Palantla Chinantec: oral ~ lightly nasalized ~ heavily nasalized, expressed
in terms of features: [–nasal, +oral] ~ [–nasal, –oral] ~ [+nasal, –oral].

Rather than indicating the asymmetry in the assimilatory behavior of [oral] by
simple formal elimination of this feature, it seems preferable to 1) focus additionally on
understanding the phonetic mechanisms involved and 2) include the observed featural
restrictions in a cross-linguistic typology of phonological tendencies. It may be of
interest, for example, to determine whether the relative activation or displacement of the
[nasal] agonist muscle group is substantially greater than the activation or displacement
of the [oral] agonist muscle group. If so, the difference might account for the
asymmetry in assimilation behavior.

Returning to the issue of voice assimilation, the primary motor dimension of
position is characterized by opposing action between an extreme vocal fold position
(glottal noise) and a central one (pure voice), as discussed in Section 2.4.2.3. Thus
[voice] and [noise] form an equipollent pair. As for the secondary motor dimension of
striction, an even clearer picture emerges of the antagonistic action between constricted
and spread vocal folds. Hence [constricted] and [spread] constitute another equipollent
pair.

Frequent voiceless-to-voice assimilation is most likely due to the spread of
agonist-excitatory, positively specified [+voice]. The voice-to-voiceless assimilation seen
in Yorkshire English, Parisian French, Ya:thê, and Makkan Arabic probably occurs less
frequently because it involves the spread of the trigger set [–voice, +spread], which contains agonist-inhibitory [–voice]. Recall that the agonist-inhibitory feature –F is presumably not as phonologically active as the agonist-excitatory feature +F (cf. Steriade’s remarks above on the spread of [–high] in Bantu).

7.2.2.2 Natural classes of phonation types

The features in Table 7.10 permit the characterization of the following natural classes of glottal manner:

A) [+voice]. The natural class comprising pure voice [+voice, –noise] and noisy voice [+voice, +noise]. The three types of noisy voice (harsh, harsh whispery, and breathy voice) are used as allophones of pure voice in the numerous languages without lexically contrastive pure voice and noisy voice. For example, the lexical meaning of a word does not change when pronounced with harsh voice, only the paralexical meaning does (e.g. in anger).

B) [+noise]. The natural class comprising glottal noise [–voice, +noise] and noisy voice [+voice, +noise]. There are languages that treat /ʔ/ and voicelessness (breath and whisper) as belonging to the natural class of glottal noise [–voice] (see Section 2.4.2.1 as well as White Tai and Nung below). Because a number of languages have noisy voice allophones of /ʔ/ and /h/, i.e. harsh and breathy voice respectively, it is possible that [+noise] constitutes a natural class in those languages (see Gordon and Ladefoged, 2001: 391-392).

C) [–voice]. The natural class of glottal noise (glottal stop, breath, and whisper; see Section 2.4.2.1 as well as White Tai and Nung below).
D) [-noise]. The natural class of pure voice in opposition to either harsh or breathy voice (see Gujarati and Jalapa Mazatec in Chapter 6).

E) [+constricted]. The natural class including glottal stop, whisper, harsh voice, harsh whispery voice, tense voice, plain voice (see Siamese and Songkhla below).

F) [+spread]. The natural class including whisper, breath, harsh whispery voice, breathy voice, plain voice, lax voice.

G) [-constricted]. The natural class including breath, breathy voice, and lax voice; the three phonation types may all be allophones of one another, and of /h/ in particular (cf. Gujarati, where /h/ > breathy voice, then breathy voice > lax voice in Chapter 6).

H) [-spread]. The natural class including glottal stop, harsh voice, and tense voice; the three phonation types may all be allophones of one another, and of /ʔ/ in particular (cf. Jalapa Mazatec, where historically /ʔ/ > harsh voice, then harsh voice > tense voice in Chapter 6).

I) [-voice, +spread]. The natural class of voicelessness, including whisper and breath (see Section 2.4.1 and Lung Ming below).

Weidert and Subba (1985: 24) give the following examples of the intervocalic and post-nasal voiceless-to-voice rule in the Tibeto-Burman language Limbu:

a) p → b

b) ph → bfi

c) t → d

d) th → dfi
e) \( c \rightarrow dz \)

f) \( k \rightarrow g \)

g) \( kh \rightarrow g\tilde{h} \)

The voiceless aspiration subsegment \([h]\) in \([p][h]\) differs from the voiced aspiration subsegment \([\tilde{h}]\) in \([b][\tilde{h}]\) only in the feature \([\pm\text{voice}]:\)

1) \([h]\) (breath): \([-\text{voice}, +\text{noise}, –\text{constricted}, +\text{spread}]\)

2) \([\tilde{h}]\) (breathy voice): \([+\text{voice}, +\text{noise}, –\text{constricted}, +\text{spread}]\)

Observe the ease with which the voiceless and voiced aspirated sounds pattern together, probably because breath and breathy voice share all but one feature value (see Lombardi, 1994: 7-10 on voiceless and voiced aspiration).

In Chapter 2, the phonological contrast between whisper and breath was established by means of evidence from Gandour’s 1975 study of Proto-Tai. With a view toward natural classes, it is useful to re-examine the categories of onset consonants that condition the three tones of Proto-Tai. His categories of onset consonants are:

I. voiceless aspirated stops and \( h \), symbolized by \( \text{ph} \): aspirated sounds with breath \([-\text{voice}, –\text{constricted}]\)

II. voiceless fricatives and voiceless sonorants, symbolized by \( f \): unaspirated sounds with breath \([-\text{voice}, –\text{constricted}]\)

III. voiceless unaspirated stops, symbolized by \( p \): unaspirated sounds with whisper \([-\text{voice}, +\text{constricted}, +\text{spread}]\)

IV. glottal stop, preglottalized stops, and preglottalized glides, symbolized by \( ? \): glottal stop \([-\text{voice}, +\text{noise}, –\text{spread}]\) and harsh voice \([+\text{voice}, +\text{noise}, –\text{spread}]\)
V. voiced obstruents and sonorants, symbolized by $b$: voice [+voice]

The categories of onset consonants select which reflexes of the Proto-Tai tones follow, patterning together or separately depending on the dialect:

A) White Tai ($\text{ph, f, p, } ?$) vs. ($b$): a two-way contrast between 1) [–voice] (glottal noise, neglecting the harsh voice of $?$) and 2) [+voice].

B) Lung Ming ($\text{ph, f, p}$) vs. ($?, b$): a two-way contrast between 1) [–voice, +spread] (voicelessness) and 2) [+voice] (neglecting [?] of $?$).

C) Siamese ($\text{ph, f}$) vs. ($p, ?, b$): a two-way contrast between 1) [–constricted] (breath) and 2) [+constricted] (whisper, [?] and harsh voice, voice if constricted).

D) Songkhla ($\text{ph, f}$) vs. ($p, ?, b$) vs. ($b$): a three-way contrast among 1) [–voice, –constricted] (breath), 2) [–voice, +constricted] (whisper and [?] if $?$ is [?]), and 3) [+voice].

E) Nung (Bac Va) ($\text{ph}$) vs. ($f, p, ?$) vs. ($b$): a three-way contrast among 1) aspirated [–voice, –constricted] (breath), 2) unaspirated [–voice] (glottal noise if $?$ is [?]), and 3) [+voice].

Compared with current systems of glottal manner features such as those proposed by Kenstowicz (Table 7.5) and Lombardi (Table 7.6), the present framework (Table 7.10) provides a more comprehensive and detailed account of the natural classes of phonation type in addition to furnishing a measurable acoustic correlate for each feature (Tables 7.8-7.9). For example, note the parallelism of the contrasts between the constricted sounds /$?$/ (glottal stop) ~ $V$ (harsh voice) in the $?$ class and the Limbu aspirated
subsegments [h] (breath) ~ [ɦ] (breathy voice). In both cases, only [–voice] changes to
[+voice], the other features retain their original values. Although Kenstowicz’s [voice]
and [constricted, spread] features could account for these oppositions, it is only at the cost
of confounding harsh voice with tense voice and breathy voice with lax voice, given that
his system lacks a [noise] feature distinguishing noisy voice from pure voice.
Furthermore, no previous binary-based feature system can distinguish between breath and
whisper, a required contrast in C) Siamese and D) Songkhla. Finally, the category of
glottal noise (glottal stop, whisper, breath) is likewise unattainable in other classificatory
systems, but needed in A) White Tai and E) Nung (see also Section 2.4.2.1).

7.2.2.3 Equipollent perceptual and gestural markedness

In the course of the discussion on phonetic scales, it was pointed out that two- or
three-way divisions of the perceptual continuum occur quite frequently (Section 7.1.6).
As antagonistic action often similarly gives rise to three-way contrasts in a motor
dimension, it may be useful to explore the markedness implications of a simple
equipollent opposition with three terms (see Hayes and Steriade, 2004 for a current
discussion of the notion of markedness).

In their investigation of universal tendencies in consonant systems, Lindblom and
Maddieson (1988: 72) make the following generalization: “Consonant inventories tend to
evolve so as to achieve maximal perceptual distinctiveness at minimum articulatory
cost.” A number of studies have elaborated the hypothesis of a trade-off between
perceptual distinctiveness and gestural economy (Lindblom, 1983, 1986; Boersma, 1998;
Kirchner, 2001; Flemming, 2002, 2004). In an attempt to formalize the conflicting
requirements of distinctiveness and gestural economy, Flemming (2002: 4) poses three general constraints, two of which are relevant here:

I. Maximize the distinctiveness of contrasts

II. Minimize articulatory effort

To determine the markedness relations of the three-term equipollent opposition, let us examine the equipollent pair of features \([F_1, F_2]\) with the intermediate term characterized by two negatively specified features \([-F_1, -F_2]\), as shown in Table 7.12.1. This first type of equipollent feature structure is the same as that of the vowel height pair [high, low]. Observe that the agonist-excitatory, positively specified \(+F_1\) and \(+F_2\) are gesturally marked because both are in violation of constraint II, i.e. neither minimizes articulatory effort. However, as \(+F_1\) and \(+F_2\) are maximally distant from each other on the perceptual scale, they are perceptually unmarked since both obey constraint I, i.e. each maximizes the distinctiveness of contrasts.

Now let us examine the equipollent pair of features \([F_1, F_2]\) with a middle term characterized by two positively specified features \([+F_1, +F_2]\), as in Table 7.12.2. This second type of equipollent feature structure is the same as those of the glottal position feature pair [voice, noise] and the glottal stricture feature pair [constricted, spread]. The only difference in the markedness relations between the negatively and positively
specified middle terms is that \([-F_1, -F_2]\) is perceptually marked, whereas \([+F_1, +F_2]\) is both gesturally and perceptually marked. The positively specified middle term \([+F_1, +F_2]\) is not only equidistant from each of the two extreme perceptual regions and therefore violates constraint I (like \([-F_1, -F_2]\)), but it is also agonist-excitatory and hence violates constraint II (unlike \([-F_1, -F_2]\)).

Since \([+F_1, +F_2]\) is doubly marked, it should be strongly avoided in comparison to \([-F_1, -F_2]\). This indeed appears to be the case. Noisy voice, the intermediate term of glottal position \([+\text{voice}, +\text{noise}]\), occurs far less often than the mid vowel, the intermediate term of vowel height \([-\text{high}, -\text{low}]\). The phonation types of UPSID sonorants (nasals, lateral liquids, trilled rhotics, glides) were discussed in Section 2.4.2.2.

The absolute frequencies of the sonorant phonation types are:

1) Pure voice \([+\text{voice}, -\text{noise}]\): 1854

2) Noisy voice \([+\text{voice}, +\text{noise}]\): 74 (70 laryngealized and 4 breathy)

Maddieson (1984: 124) gives the absolute frequencies of the three degrees of vowel height and backness as follows:

1) High vowel \([+\text{high}, -\text{low}]\): 994

2) Mid vowel \([-\text{high}, -\text{low}]\): 1032

3) Low vowel \([-\text{high}, +\text{low}]\): 523

Table 7.12.2 Gestural and perceptual markedness of the equipollent feature pair \([F_1, F_2]\) with a middle term characterized by two positively specified features \([+F_1, +F_2]\), like the glottal position and stricture feature pairs ([voice, noise], [constricted, spread])
1) Front vowel [+front, –back]: 1019
2) Central vowel [–front, –back]: 566
3) Back vowel [–front, +back]: 964

Clearly, positively specified noisy voice [+voice, +noise] is very much avoided when compared to the negatively specified mid vowel [–high, –low] and the negatively specified central vowel [–front, –back].

The intermediate term of glottal stricture [+constricted, +spread] is difficult to evaluate for lack of published reports on the frequency of occurrence of whisper, harsh whispery voice, and plain voice. Although not rare allophonically (cf. Section 6.4), the known infrequency of lexically contrastive harsh whispery voice is mostly likely due to its highly marked feature composition: [+voice, +noise, +constricted, +spread]. As the proposed markedness relations seem to be consistent with the frequencies observed thus far, plain voice [+constricted, +spread] should occur less often than either tense voice [+constricted, –spread] or lax voice [–constricted, +spread]. Obviously experimental work is needed to confirm this prediction.

Glottal manner appears to be unique among phonological features to possess positively specified intermediate terms like the glottal position pair [+voice, +noise] and the glottal stricture pair [+constricted, +spread]. Taking the markedness relations into account, it is almost certain that the other phonological features have a negatively specified intermediate term because the middle term [–F1, –F2] occurs about as frequently as any one of the extreme terms [+F1, –F2] and [–F1, +F2]. Even tone, likewise produced with the vocal folds, appears to have a negatively specified intermediate term [–F1, –F2]. Maddieson (1978: 338) states that “languages with three tone levels are commonplace,
while those with only two are the most universally encountered type of tone language.” Yip (2002: 26) mentions that “true underlying three-tone systems are…very common, and found in all regions, such as Huajuapan Mixtec, Nupe, and Kunama.” The Halle and Stevens system of pitch features represents the three level tones as follows: high tone [+stiff vocal folds, –slack vocal folds], mid tone [–stiff vocal folds, –slack vocal folds], low tone [–stiff vocal folds, +slack vocal folds] (Table 7.2). The question then arises as to why glottal manner features have a positively specified middle term [+F₁, +F₂] rather than the usual negatively specified one [–F₁, –F₂], in which both F₁ and F₂ are agonist-inhibitory.

The evident explanation for the positively specified middle term of position [+voice, +noise] relates to the simultaneous implementation of the periodically vibrating vocal folds and aperiodicity. The feature specification of noisy voice [+voice, +noise], or more precisely [+central, +extreme], indicates the maintenance of vibration even when the vocal folds are subject to additional adduction (harsh voice), additional abduction (breathy voice), or both at the same time (harsh whispery voice). The increase in motor activity relative to pure voice [+central, –extreme] is signaled by the positively specified feature [+extreme] (or [+noise]).

The explanation for the positively specified middle term of stricture [+constricted, +spread] lies in the fact that the vocal folds are capable of simultaneous opposing movements, unlike the other speech organs that tend to move as units. For instance, it was shown in Chapter 2 (Section 2.4.2.3) that whisper is produced when the anterior or ligamental glottis is adducted [+constricted] while the posterior or cartilaginous glottis is abducted [+spread]. Recall also Laver’s description of harsh whispery voice (1980: 136):
“...the whisper component in harsh whispery voice, say, is maintained by a much greater effort on the part of the...cricoarytenoid muscles to keep the arytenoid triangle open [+spread] against the vigorous attempt by the interarytenoid muscles to close it [+constricted].”

The middle term of glottal stricture [+constricted, +spread], where both [+constricted] and [+spread] are agonist-excitatory, consequently mirrors the simultaneous activation of the mechanically opposing laryngeal muscles.

In order to demonstrate the generality of the proposed framework of glottal manner features, the next section of this work will very briefly consider height and backness features as well as the supraglottal manner features (major class features). Although a satisfactory treatment of the complexities of each topic would require a monograph in itself, the present feature framework (primary and secondary motor features, linear acoustic scale) is given more weight if it can be shown to accommodate other speech parameters. An even more important reason for examining height, backness, and supraglottal manner is to illustrate directly how exceptional glottal manner is when compared to other feature subsystems. Only glottal manner is characterized by positively specified middle terms, i.e. primary motor [+voice, +noise] and secondary motor [+constricted, +spread], while height, backness, and supraglottal manner have negatively specified middle terms, i.e. primary motor [−F₁, −F₂] and secondary motor [−F’₁, −F’₂].
The feature framework and other speech parameters

7.3.1 Height features

For comparison, the *Handbook of the International Phonetic Association* (1999) distinguishes seven vowel heights: 1) close, 2) near-close, 3) close-mid, 4) mid, 5) open-mid, 6) near-open, 7) open. Maddieson (1984: 204) uses the same seven-category system for transcription purposes: 1) high, 2) lowered high, 3) higher mid, 4) mid, 5) lower mid, 6) raised low, 7) low. Both he and the IPA assume a two-way division of the high and low vowels, and a three-way division of the mid vowel.

The height features are presented in Table 7.13. The only difference between the IPA classification and the present one consists of the addition of raised high and lowered low, following Maddieson’s terminology (1984: 204). Although nine distinct vowel heights are clearly in excess of the maximum of five (or possibly six?) phonemic height contrasts attested cross-linguistically (Crothers, 1978: 119; Disner, 1983), a feature

<table>
<thead>
<tr>
<th>F1</th>
<th>Height</th>
<th>High</th>
<th>Low</th>
<th>Raised</th>
<th>Lowered</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Raised high</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>Plain high</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>Lowered high</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>4</td>
<td>Raised mid</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>Plain mid</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>Lowered mid</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>7</td>
<td>Raised low</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>Plain low</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>Lowered low</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

*Table 7.13* The nine heights cross-classified by the two equipollent motor position features [high, low], the two equipollent motor displacement features [raised, lowered], and the F1 linear acoustic scale (1 ≤ F1 ≤ 9, where 1 is the lowest log frequency, and 9 the highest).
system must not only be classificatory, but also phonetically accurate enough to account for subphonemic variants by means of an allophonic rule of the form /x/ → [x], where x is a feature matrix (see Khabanyane, 1991 in this regard). The two equipollent motor position features [high, low] are uncontroversial, but the two equipollent motor displacement features [raised] and [lowered] represent a terminological shift, the most frequent designations being ATR (advanced tongue root) and RTR (retracted tongue root), respectively. Remarking the inadequacy of the traditional tense/lax distinction to account for Akan cross-height harmony, Stewart (1967) proposed ±ATR instead (see van der Hulst and van de Weijer, 1995 for a review of ATR/RTR harmony). However, the X-ray studies of Lindau et al. (1972) show that the African cross-height harmony in question can be implemented phonetically in several ways, including via differences in tongue height in Ateso:

“...The African languages under consideration exhibit vowel harmonies that are structurally similar to each other...but the phonetic interpretation will not be the same in the different languages. In Twi and Dho-Luo the articulatory mechanism is Advanced Tongue Root, in Ateso it is Tongue Height, and in Igbo it is partly Advanced Tongue Root, partly Tongue Height.”

As the current use of ATR/RTR is “motivated only by the need to describe a fourth height,” according to Clements and Hume (1995: 283), the more general labels [raised] and [lowered] seem to be indicated.

The F1 linear acoustic scale is needed to describe phonological processes in which the height continuum appears to be divided into a series of integral steps. Lass (1984: 105) describes a Middle English sound change (14th century) where a short vowel in an open syllable lengthened, and was lowered if it was non-low. The initial system is:
The sound changes are:

1) /i/ → [e:]

2) /e/ → [e:]

3) /u/ → [ɔ:]

4) /o/ → [ɔ:]

5) /a/ → [a:]

The part of the rule that involves vowel lowering by one step can be then stated:

\[ F1(n) \to F1(n + 1), \text{ where } 1 \leq n \leq 2 (\text{Lass, 1984: 106}). \]

Middle English vowel lowering is an illustration of a (relatively) unconditioned scalar sound change (cf. also Lindau, 1978 for Scanian). Partial height harmony is an example of a conditioned scalar sound change. Parkinson (1996: 37-38) defines partial height harmony as a process in which “a vowel moves toward another vowel but does not surface with the same height as the vowel to which it assimilates…the target surfaces with a height between its original height and that of the trigger”. He then (1996: 39-40) describes the raising of nonhigh vowels in Lena Spanish, whereby

\[
\begin{array}{c}
F1(1) \\
F1(2) \\
F1(3)
\end{array}
\begin{array}{cccc}
i & u & e & o \\
\hline
\end{array}
\]

1) a → e/\_u

2) e, o → i, u/\_u,
or alternatively, one step is subtracted from the nonhigh vowels when followed by the high vowel trigger F1(1):

\[ F1(n) \rightarrow F1(n-1)/F1(1), \text{ where } 2 \leq n \leq 3 \]

(see Contreras, 1969 for an analysis very similar to the present one).

Salting (1998, Chapter 2) observes that even closely related dialects (Italian) or languages (Bantu) diverge in their preferences with respect to the type of height organization they use. Thus some languages adhere to the motor hierarchy, i.e. those with cross-height [raised, lowered] harmony, while others tend ignore it and follow the F1 linear acoustic scale.

### 7.3.2 Backness features

The backness features are illustrated in Table 7.14. The *Handbook of the International Phonetic Association* (1999) provides for five degrees of backness: 1) front, 2) near-front, 3) central, 4) near-back, and 5) back; Crothers (1978: 119) indicates that

<table>
<thead>
<tr>
<th>F2</th>
<th>Backness</th>
<th>Front</th>
<th>Back</th>
<th>Advanced</th>
<th>Retracted</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Retracted back</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>2</td>
<td>Plain back</td>
<td>–</td>
<td>+</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>3</td>
<td>Advanced back</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>4</td>
<td>Retracted central</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>5</td>
<td>Plain central</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>6</td>
<td>Advanced central</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>7</td>
<td>Retracted front</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>8</td>
<td>Plain front</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>9</td>
<td>Advanced front</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

**Table 7.14** The nine degrees of backness cross-classified by the two equipollent motor position features [front, back], the two equipollent motor displacement features [advanced, retracted], and the F2 linear acoustic scale (1 ≤ F2 ≤ 9, where 1 is the lowest log frequency, and 9 the highest).
the maximum number of distinctions for backness is four. Lindau (1978) points out that three phonological degrees of backness are necessary to account for Norwegian front rounded, central rounded, and back rounded high vowels (see also Campbell, 1974 on this point). Hence the adoption of the primary motor position pair [front, back] is clearly warranted.

Lass (1984: 107) finds no evidence in support of the F2 linear acoustic scale, but states that the option should be left open.

7.3.3 Supraglottal manner features (major class features)

Nasal consonants (e.g. m, n, ƞ) are not included among the supraglottal manner types shown in Table 7.15 because doing so would entail a discussion that goes far beyond the scope of this brief overview. It is useful to compare the above arrangement of supraglottal manner types with the sonority hierarchies proposed by Hankamer and Aissen (1974) as well as Selkirk (1984).
Hankamer and Aissen (1974: 132) order the supraglottal manner types as follows:

1) stops
2) s
3) nasals
4) l
5) v (pronounced w)
6) j
7) r
8) vowels

They base their ranking on assimilation in Pali: the consonant lower in the series (toward 8) completely assimilates to the consonant higher in the series (toward 1).

Selkirk (1984: 112) offers a somewhat different ranking of supraglottal manner:

1) p, t, k
2) b, d, g
3) f, θ
4) v, z, ð
5) s
6) m, n
7) l
8) r
9) i, u and their approximant counterparts j, w
10) e, o
11) a
Observe that Selkirk places the glides j, w after approximant r—the most widely accepted ranking—whereas Hankamer and Aissen find the opposite order in Pali assimilation phenomena. Consequently, there may be language-specific variation on how glides and r are ordered within the natural class of central approximants [–open, –close, –narrowed].

Stops, e.g. [t, d], and the nonstrident fricatives, e.g. [θ, δ], are assumed to form the natural class [–open, +close, –widened] on account of the non-occurrence of the English onset sequences *tl, *θl, *dl, *δl (Clements and Keyser, 1983: 44; cf. also Hall. 1997: 36 on this restriction) as well as the frequent substitution of alveolar stops for interdental fricatives in child and second language acquisition. Further evidence that nonstrident fricatives constitute a category intermediate between stops and strident fricatives may be their somewhat rarer occurrence. Because nonstrident fricatives are equidistant from each of the two extreme obstruent types, i.e. stops and strident fricatives, they could be treated as especially perceptually marked and thus violate constraint I in Section 7.2.2.3, i.e. they do not maximize the distinctiveness of contrasts. Finally, in Lavoie’s sample of lenition processes (2001: 34-35), a voiced bilabial stop [b] weakens to a nonstrident voiced fricative [β] (12/17) substantially more often than it does to a strident fricative [v] (5/17). If the supraglottal stricture of nonstrident fricatives is indeed intermediate between those of stops and strident fricatives, this would be the expected outcome. Her results for voiceless stops and the other places of articulation are less clear. Note, however, that there are no instances where [d] → [z], but several in which [d] → [ð].

Important natural classes of supraglottal manner are:
A) [+open]. Vowels.

B) [+close]. Obstruents.

C) [--open, --close]. Approximants.

D) [--open]. Consonants (obstruents and approximants inclusive of glides). For instance, *an, the external sandhi form of the English indefinite article occurs before consonants, whereas *a occurs before vowels, *an thought, *an laugh, *an rabbit, *an yacht, *an walk.

E) [--close]. Sonorants (vowels and approximants inclusive of glides).

F) [--open, +close, --narrowed]. Fricatives (strident and nonstrident fricatives).

G) [--open, --close, --widened]. Liquids (the alveolar lateral approximant l and the narrow central approximant when the latter is r).

H) [--open, +narrowed] Narrowed harmony of stop and lateral approximant consonants, both of which frequently pattern together. For example, consider the non-weakened Spanish stop in cal[d]o “broth”, but the weakened stop in ver[ð]e “green” (Kenstowicz, 1994: 35-36; Mielke 2004: 246 found in his cross-language survey that lateral liquids pattern 55% of the time with continuants, 45% of the time with noncontinuants).

To characterize the acoustical properties of their vocalic and consonantal features, Jakobson et al. (1952: 46) propose the following definitions (cf. also Fant, 1973: 155):

“Phonemes possessing the vocalic feature are acoustically characterized by the presence of formants with small damping and hence with a relatively narrow band width.
Phonemes possessing the consonantal feature are acoustically characterized by a broadening, reduction and fusion of formants and formant regions due to zeros, high damping or transient variations of formant frequencies.”
The linear acoustic scale of the supraglottal bandwidth (SBW) shown in Table 7.15 is therefore designed to reflect the damping of the vocal tract filter, with SBW(1) indicating the greatest damping (stops), and SBW(9) the least damping (wide-stricture vowels). Note that the voicing contrasts in Selkirk’s sonority ranking are not relevant for supraglottal manner, which is purely a function of the vocal tract filter. Glottal manner is mixed in this regard, with the glottal noise contrasts (glottal stop, whisper, breath) being cued by differences in the first formant peak energy factor (P1E), itself a function of the vocal tract filter.

Parker (2002: 106) found a very high correlation of 0.96-0.97 between a typical sonority hierarchy like Selkirk’s and acoustic intensity in all his data sets. Because acoustic intensity grows with decreasing damping of the vocal tract filter, these results are a good indication that supraglottal bandwidth is the appropriate linear acoustic scale. Nevertheless, overall acoustic intensity has the disadvantage of simultaneously measuring the glottal source amplitude. Preliminary tests suggest that the second formant peak energy factor (PE2) is likewise an effective estimator of supraglottal bandwidth (SBW) in addition to being a theoretically more plausible Q-like measure (see Section 4.5).

Similarly to vowel height (F1), it is fairly easy to find phonological processes in which the SBW linear acoustic scale appears to be divided into a series of integral steps. Hankamer and Aissen (1974: 136-139) considered it problematic to describe the Pali rule of complete consonant assimilation by means of classificatory features. They put forward a scalar assimilatory rule (our notation, nasals neglected):

\[ \text{SBW}(m) \rightarrow \text{SBW}(n)/\_\_ \text{SBW}(n), \text{ where } m \geq n. \]
Selkirk (1984: 113) remarks that some natural classes, such as s, m, n, l, r (consonants appearing in the second position of Italian rimes), “have no simple expression in terms of major class features and can only be designated by a disjunction of feature complexes.” To account for such cases, she proposed a scalar treatment in which the sounds are inclusively indexed by their supraglottal bandwidth: $3 \leq \text{SBW} \leq 5$ (assuming r is the narrow central approximant; our notation, nasals neglected).

The phonetically conditioned alternation between the high vowels [i, u] and the glides [j, w] is a frequent phenomenon in the world’s languages (see Levi, 2004 for a number of examples). Yet according to the motor hierarchy, [i, u] are [+open] vowels and the glides [j, w] are [–open] consonants. Because vowels and glides fall into dichotomous categories, the alternation between [i, u] and [j, w]—occasioned by the position of /i, u/ in the syllable—seems rather unusual in light of the behavior of other vowels. If, on the other hand, the high vowel is produced with a narrow stricture SBW (7) and the glide is realized as the wide central approximant SBW(6), then vowel and glide are contiguous on the SBW linear acoustic scale. The motor hierarchy may be ignored in this particular case, resulting in the class of the vocoid $^{12}$ ($6 \leq \text{SBW} \leq 7$), i.e. phonemic /i, u/ with the alternating position-dependent allophones [i, u] and [j, w]. A similar tendency to dispense with the motor hierarchy was already seen for vowel height.

### 7.3.4 Organizing principles of the four feature subsystems

Recapitulating, there appears to be a maximum of nine possible categories in each of the feature subsystems considered above (glottal manner, supraglottal manner, height,
backness). However, the number of lexically contrastive categories never reaches nine in the individual language.

The nine categories are classified by a three-by-three motor hierarchy determined by a primary equipollent feature pair \([F_1, F_2]\) and a secondary equipollent feature pair \([F'_1, F'_2]\). The secondary equipollent features \([F'_1, F'_2]\) give rise to a maximal three-way distinction within each of the three categories established by the primary equipollent features \([F_1, F_2]\). The intermediate terms of both the primary and secondary motor features are negatively specified \([-F_1, -F_2]\) and \([-F'_1, -F'_2]\), with the one exception of the positively specified intermediate terms of glottal manner: \([+\text{voice}, +\text{noise}]\) and \([+\text{constricted}, +\text{spread}]\).

The primary equipollent feature pairs of glottal manner, supraglottal manner, height, and backness fix the \textit{position} of the speech organs (\([\text{voice, noise}], [\text{open, close}], [\text{high, low}], [\text{front, back}]\)). The secondary equipollent feature pairs of glottal manner and supraglottal manner specify the degree of \textit{stricture} within each position (\([\text{constricted, spread}], [\text{narrowed, widened}]\)) whereas the secondary equipollent feature pairs of height and backness control the \textit{displacement} within each position (\([\text{raised, lowered}], [\text{advanced, retracted}]\))

The nine categories may also be classified by means of a linear acoustic scale. If the acoustic scale is log formant frequency \(F\) (Section 4.2.1), then it can be represented as \(1 \leq F \leq 9\), where 1 is the lowest log frequency, and 9 the highest, such as the \(F1\) scale for height and the \(F2\) scale for backness. If the acoustic scale is bandwidth \(BW\), then it can be represented as \(1 \leq BW \leq 9\), where 1 is the largest bandwidth, and 9 the narrowest, like the GBW scale for glottal manner and the SBW scale for supraglottal manner.
7.4 General summary and directions for future research

This chapter begins by providing a critical overview of glottal manner feature systems. Initial attention is focused on the phonetically homogeneous subsegment, since all feature systems up to the present consider the possibly complex phonemic segment as the basic unit of analysis, making it difficult to associate a given phonological feature with its acoustic correlate. In order to address this problem, a set of criteria for determining the boundaries of the phonetically homogeneous subsegment is given in Chapter 4 (Section 4.6).

Much of Keating’s 1988 criticism of the Halle and Stevens system (1971) is directed against permanent associations between features, as between tone and voice or tone and phonation type. Later systems of glottal manner like Kenstowicz’s (1994) do not assume fixed associations between features, and it is proposed that when linkages do occur they should be expressible by means of a conditional if-then rule in the individual language or within a broader typological statement. The frequent co-occurrence of considerable modulation noise (harsh or laryngealized phonation) with low-pitch register (creaky voice) is mentioned as an example of such a conditional rule (Section 7.1.5). The phonetic and phonological inadequacies of privative feature systems are likewise discussed, the main phonological difficulty with privativity being its failure to account for recurrent natural classes defined by negatively specified members of an equipollent feature pair, such as [–high] and [–low].

Before taking into consideration the proposed glottal manner features, the nine phonation types and their acoustic cues are presented following the motor-acoustic
classification at the end of Chapter 2. The production mechanisms of glottal parameters such as modulation noise (harsh voice), aspiration noise (breathy voice), and spectral tilt are treated at length in Chapter 3. The topics of the subsequent chapters include the signal processing techniques developed to estimate the glottal parameters (Chapter 4) as well as the evaluation of these techniques using synthesized waveforms (Chapter 5) and natural exemplars of the nine phonation types (Chapter 6). Hence the acoustic correlates are confirmed experimentally prior to the feature decomposition of the phonation types.

The motoric interpretation of phonological features put forth by Halle (1983) is extended in this work. Opposing muscle sets are assumed to be actuated by a pair of equipollent features \([F_1, F_2]\), in which the positively specified feature \([+F]\) excites the agonists (and inhibits the antagonists) while the negatively specified feature \([-F]\) inhibits the agonists (and excites the antagonists). If, in a given language, one positively specified member of the equipollent pair \(+F_1\) triggers assimilation but the other \(+F_2\) does not, it is suggested that this asymmetry may be due to the greater activation or displacement of the muscle group commanded by \(+F_1\), a phonetically verifiable hypothesis (Section 7.2.2.1).

In Table 7.10 the nine phonation types are classified by a pair of primary motor position features [voice, noise] (or equivalently [central, extreme]), a pair of secondary motor stricture features [constricted, spread], as well as a linear acoustic scale of derived glottal bandwidth \(1 \leq GBW \leq 9\), where 1 is the largest glottal bandwidth, and 9 the narrowest (Section 2.4.2.4). The present feature framework captures natural classes of glottal manner unattainable in other binary-based classificatory systems (e.g. Halle and Stevens, 1971, Kenstowicz, 1994), one instance being the category of glottal noise (glottal stop, whisper, breath) in White Tai and Nung (Section 7.2.2.2). In addition, the
analysis of the perceptual and gestural markedness of the simple three-term equipollent opposition helps to clarify why phonemic harsh whispery voice seldom if ever occurs (Section 7.2.2.3). Lastly, the feature subsystems of height, backness, and supraglottal manner are demonstrated to share the same organizing principles as glottal manner, with the sole exception that only glottal manner possesses the positively specified intermediate terms [+voice, +noise] and [+constricted, +spread], all other feature subsystems have negatively specified ones [–F1, –F2], and [–F´1, –F´2] (Section 7.3.4).

Among the future directions of this research, more phonetic and phonological evidence needs to be provided in support of the height, backness, and supraglottal manner feature subsystems outlined in Sections 7.3.1-7.3.3. For supraglottal manner in particular, formal tests are required to assess how well the second formant peak energy factor PE2 estimates the supraglottal bandwidth SBW.\textsuperscript{14}

As discussed in Chapter 2 (Section 2.4.2.4), the typological data on phonation types clearly substantiate the idea of a linear acoustic scale of glottal bandwidth GBW, the extrema of which are

\begin{itemize}
  \item GBW(1): the very broad GBW of the pulse-like attack or release of a glottal stop,
  \item GBW(9): the very narrow GBW of pure-tone-like lax voice.
\end{itemize}

A more direct approach to validating the nine-step GBW scale shown in Table 7.9 would be a study in which one or more experienced phoneticians produce the nine phonation types. A bandwidth measure such as the second spectral moment could then be computed over the entire spectrum and compared against the GBW linear scale (see Forrest et al., 1988 on the calculation of spectral moments).
In Section 2.4.2.3 of Chapter 2, it was mentioned that there appear to be no electromyographic studies that systematically explore the different phonation types using the same method and participants. To ensure uniformity of the data collected, a phonetician could produce the nine phonation types while the electrical activity of the laryngeal muscles is recorded. As a final point, although several speech waveforms in Chapter 6 exemplify rather well each of the nine phonation types, a larger, more systematic sample of languages and speakers would be useful for confirming the results obtained thus far.
NOTES TO CHAPTER 7

1. “Wenn man Beach (1938) folgt, so gibt es z. B. im Korana-Dialekt des Hottentottischen eine laterale Click-Affrikata || mit einem aspirierten velaren affricativen Absatz, d. i. ein fünfkolumniges Segment: Verschluss + || + k + x + h.” The symbol || replaces Sasse’s older notation for the lateral click. See Ladefoged and Maddieson (1996, Chapter 8) for a recent treatment of clicks and their accompaniments.

2. The uvular ejective of Quechua q’aña ‘tomato sauce’ in Figure 6.3.2 provides a good illustration of a long glottal closure with a strong release.

3. The phonemic contrast between Korean lenis and aspirated stops may be expressed as /[k][fi]/ vs. /[k][h]/ in IPA notation, assuming that length of aspiration is the only difference between them. The IPA extra-short diacritic şe indicates the relatively brief 35 ms duration of the [h] in the lenis stop in comparison to the more typical 90 ms speech sound duration of the [h] in the aspirated stop (see Kim 1970 for the VOT values of 35 ms and 90 ms).

4. For an analogous copying of a laryngeal subsegment to another phoneme, consider one of the examples of Bartholomae’s Law provided by Szemerényi (1989: 107): Sanskrit root labh- ‘grasp’ + ta → verbal adjective labdha ‘grasped’. The breathy voice subsegment [fi] of the complex segment /[b][fi]/ spreads to the voice assimilated /d/,

resulting in the complex segment /[d][fi]/ (see also Bartholomae’s Law in Collinge, 1985, Ladefoged and Maddieson 1996: 69-70 on aspirated voiced stops, and Note 7 below).
5. With regard to the pitch features [stiff] and [slack], Yip (2002: 56-59) provides a critical discussion of current successors to the Halle and Stevens system.

6. Traill (1985: 69) mentions the major co-occurrence restriction that pharyngealized vowels can not be [–back].

7. The IPA “voiced glottal fricative” symbol [fi] “represents a breathy voiced sound, rather than an ordinary voiced sound,” according to the Handbook of the International Phonetic Association (1999: 20). One suggestion for symbolizing harsh whispery voice is to use [fi] in conjunction with the IPA diacritic for tense or harsh voice: [fi]. A further use of the tense or harsh diacritic would be to indicate whisper by [h] and breath by plain [h], as whisper is produced by the combination of a constricted and spread glottis (see Section 2.4.2.3).

8. The long tradition of abstract and minimalist approaches to phonological representations may be a consequence of alphabetic writing (Port, to appear). To indulge in some speculation, one of the most influential figures in modern phonology, the Russian Slavist N. S. Trubetzkoy, was certainly familiar with the phonological analysis implicit in the Classical Greek opposition between the symbols θ and τ. Yet these unanalyzable symbols tend to relegate to the background the observation that θ is a complex segment. For this reason, he may not have attached a great deal of importance to the fact that utterance-initial θ is made up of two phonetically homogeneous subsegments /tᵣ/[h] (the r subscript indicating the release phase of the stop), while utterance-initial τ consists of just one subsegment /t/₁. The Devanagari Indian script likewise contrasts the unanalyzable symbols थ tʰa and द ta. This may have further encouraged Trubetzkoy to
treat the opposition between Sindhi aspirated $t^h$ and unaspirated $t$ as a manner correlation parallel to the one between voiceless $t$ and voiced $d$ (Trubetzkoy, 1967: 149). In a similar vein, it is of interest to point out that two-subsegment affricates are represented as unanalyzable symbols in the Cyrillic alphabet (e.g. Russian Ц /[t][s]/ vs. Т /[t]/). Note, however, that another important figure in the history of modern phonology, the American linguist C. F. Hockett, was quite willing to adopt a cluster analysis if the phonemic patterning seemed to warrant it. For example, he states (1955: 107):

“Sanskrit, and certain modern languages such as Hindustani, are often said to have four types of stops: voiceless and voiced, intersecting unaspirated and aspirated. But in both named cases the aspiration (be it voiceless or voiced) is rather patently simply the phoneme /h/, which recurs elsewhere: this leaves just a two-way manner contrast.”

The broad interpretation of the feature [voice] is perhaps another instance of excessive abstraction in phonology. The acoustic correlates of broadly interpreted [voice] may include not only speech wave periodicity, but also stop voicelessness followed by either no voice onset time (VOT) or one of very brief duration (see van Rooy and Wissing, 2001 for detailed discussion and references). Instead of attempting to extend the compass of the feature [voice], it seems preferable to treat subphonemic variants as either freely-varying or context-dependent allophones of the same phoneme—the traditional approach. To illustrate, consider three utterance-initial allophones of the English phoneme /d/. The fully voiced allophone may be represented subsegmentally as $[[d]_c[d]_c]$, the partially devoiced allophone as $[[d]_c[t]_c[t]_c]$, and the voiceless allophone with no VOT as $[[t]_c]$, where the $c$ subscript indicates the closure phase of the stop. By contrast, the allophone $[[t]_c[h]]$ is the typical utterance-initial realization of the English phoneme /t/.
Remark that a fine-grained phonetic representation does not entail a loss of economy in the phonological description, since the phonemes /d/ and /t/ remain minimally contrastive in a given environment regardless of their respective subsegmental constituents.

9. Mid vowels presented an early problem for binary features, which led to the de facto acceptance of the equipollent pair [high, low], each member having positively and negatively specified values. Halle (1957: 71) notes that:

“These two extremes of vocal tract shape, the horn and the Helmholtz resonator, are taken as the defining characteristics of the features compact ~ noncompact [i.e. +low ~ –low] (horn shape or not) and diffuse ~ nondiffuse [i.e. +high ~ –high] (Helmholtz resonator shape or not). In terms of these features the close vowels are characterized as diffuse [+high] and noncompact [–low]; the open vowels like [a] and [æ] are classified as compact [+low] and nondiffuse [–high]; and vowels of intermediate degree of openness—e.g. [e] [o] and [ɔ]—are both noncompact [–low] and nondiffuse [–high].

... Only in the case of the feature diffuse ~ nondiffuse [+high ~ –high] has the insistence upon binary features led us to introduce a parameter which has an extremely restricted applicability and therefore may be said not to be optimal. It is for this reason that in previous formulations of the distinctive feature framework the feature compact ~ noncompact [+low ~ –low] was defined as a ternary feature. In recent months we have been led to accept the more consistent solution of postulating two binary features in place of the ternary one, because, in connection with our work on evaluation procedures for alternative phonemic solutions, we found that the consistently binary system fitted our requirements better than the mixed system previously used.”

10. See also Ladefoged, 1971: 34-35. Ladefoged and Maddieson (1996: 299-300), propose a diphthongal interpretation of the three-way vowel nasalization in Palantla Chinantec. The nasal airflows measured by Merrifield and Edmondson (1999) confirm the time-varying character of P. Chinantec nasalization, although three degrees of nasality do seem to occur during the mid portion of the vowel (Figs. 1-3).

11. A similar question may be asked in regard to the equipollent pair [round, drawn] of the more accessible lips.
12. Following Pike (1943), Clements (1990: 293) considers the vocoid to be “simply the converse of the traditional feature ‘consonantal’...” According to Chomsky and Halle (1968: 68), the feature [–consonantal] defines the class of vowels and glides.

13. Note also that there has been a recent shift away from the hypothesis of universally fixed dependency relations among feature subsystems (e.g. the Feature Geometries of Clements, 1985 and Sagey, 1986) toward the recognition that feature subsystems can be variably grouped together in language-specific ways (Padgett, 2000 and Yip, 2005, as well as Halle et al., 2000 to some extent).

14. Once the supraglottal bandwidth SBW and its associated linear acoustic scale \(1 \leq \text{SBW} \leq 9\) are better established, it may be useful to examine the concept of \textit{harmony} between the GBW of the glottal source and the SBW of the vocal tract filter. To cite one example, Lavoie (2001: 38-39) observes that the voiceless plosives \(\text{[p, t, k]}\) typically weaken to \(\text{[?]}\), whereas the voiceless fricatives \(\text{[s]}\) and especially \(\text{[x]}\) often give \(\text{[h]}\) (see also the discussion relative to Bessell, 1992 in Section 2.4.2.1). The latter process seems quite reasonable since the glottal voicelessness \(\text{[h]}\) realized simultaneously with \(\text{[s, x]}\) does not change when the supraglottal aperture increases from \([+\text{close}]\) to \([+\text{open}]\) (“debuccalization”).

In contrast, the glottal voicelessness \(\text{[h]}\) concurrent with the supraglottal plosives \(\text{[p, t, k]}\) becomes a glottal stop \(\text{[?]}\) when \([+\text{close}] \rightarrow [+\text{open}]\). The most likely explanation for this behavior is that the GBW of \(\text{[h]}\) \(2 \leq \text{GBW} \leq 3\) \textit{harmonizes} with the SBW of the original supraglottal stop (SBW = 1) to yield a glottal stop \(\text{[?]}\) with the same bandwidth rank (GBW = 1).

272
APPENDIX A

ANALYSIS OF THE MPI SPEECH SAMPLES
## Table A.1 Duration, glottal source, and formant parameters of plain voice high level si ‘four’ in Mpi.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mpi: man</td>
<td>plain high level si ‘four’ 1st subseg.* 330.7-798.5: 467.8 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>146.82</td>
<td>157.81</td>
<td>160.11</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>-0.8277</td>
<td>-0.7771</td>
<td>-0.8198</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>23.330</td>
<td>23.330</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>-4.092</td>
<td>-4.657</td>
<td>-4.657</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>-12.082</td>
<td>-12.753</td>
<td>-9.885</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>439.97</td>
<td>315.63</td>
<td>324.88</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>7.55783</td>
<td>6.90764</td>
<td>6.72705</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>1919.15</td>
<td>2033.27</td>
<td>2033.27</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>4.15399</td>
<td>4.76857</td>
<td>4.17641</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mpi: man</td>
<td>plain high level si ‘four’ 2nd subseg. 800.9-881.7: 80.8 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>172.10</td>
<td>164.80</td>
<td>157.81</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>-0.9013</td>
<td>-1.2109</td>
<td>-1.7021</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>23.330</td>
<td>23.330</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>-7.808</td>
<td>-13.358</td>
<td>-9.972</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>1.690</td>
<td>8.134</td>
<td>5.186</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>515.71</td>
<td>320.22</td>
<td>306.64</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>5.51154</td>
<td>3.08396</td>
<td>5.14400</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>2092.85</td>
<td>2123.29</td>
<td>1613.80</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>3.15613</td>
<td>0.53528</td>
<td>0.44303</td>
</tr>
</tbody>
</table>

## Table A.2 Duration, glottal source, and formant parameters of tense voice high level si (name) in Mpi.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mpi: man</td>
<td>tense high level si (name) 298.9-803.4: 504.5 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>157.81</td>
<td>160.11</td>
<td>167.20</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>-0.9214</td>
<td>-0.8170</td>
<td>-0.7484</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>23.330</td>
<td>23.330</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>-3.904</td>
<td>-4.374</td>
<td>-4.562</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>-12.089</td>
<td>-14.034</td>
<td>-10.807</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>466.13</td>
<td>315.63</td>
<td>334.40</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>7.43679</td>
<td>6.65824</td>
<td>7.09190</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>2217.29</td>
<td>2383.31</td>
<td>2315.46</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>4.47527</td>
<td>4.62629</td>
<td>4.03576</td>
</tr>
<tr>
<td>Parameters</td>
<td>Initial Measure</td>
<td>Mid Measure</td>
<td>Final Measure</td>
</tr>
<tr>
<td>---------------------</td>
<td>-----------------</td>
<td>-------------</td>
<td>---------------</td>
</tr>
<tr>
<td><strong>Mpi: man</strong></td>
<td>plain high falling si ‘die’ 284.2-788.7: 504.5 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>F0 (Hz)</strong></td>
<td>153.32</td>
<td>160.11</td>
<td>111.59</td>
</tr>
<tr>
<td><strong>Log pACC (dB)</strong></td>
<td>−0.8388</td>
<td>−0.7657</td>
<td>−1.9909</td>
</tr>
<tr>
<td><strong>HNR (dB)</strong></td>
<td>23.330</td>
<td>23.330</td>
<td>23.330</td>
</tr>
<tr>
<td><strong>Slope (dB/octave)</strong></td>
<td>−4.374</td>
<td>−4.562</td>
<td>−4.657</td>
</tr>
<tr>
<td><strong>H1–H2 (dB)</strong></td>
<td>−12.030</td>
<td>−13.447</td>
<td>−15.456</td>
</tr>
<tr>
<td><strong>F1 (Hz)</strong></td>
<td>459.45</td>
<td>320.22</td>
<td>344.20</td>
</tr>
<tr>
<td><strong>PE1 (dB)</strong></td>
<td>6.49395</td>
<td>6.81746</td>
<td>4.15841</td>
</tr>
<tr>
<td><strong>F2 (Hz)</strong></td>
<td>1975.38</td>
<td>2092.85</td>
<td>2092.85</td>
</tr>
<tr>
<td><strong>PE2 (dB)</strong></td>
<td>3.83700</td>
<td>6.82623</td>
<td>6.31504</td>
</tr>
</tbody>
</table>

**Table A.3** Duration, glottal source, and formant parameters of plain voice high falling si ‘to die’ in Mpi.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mpi: man</strong></td>
<td>tense high falling si (name) 296.4-783.8: 487.4 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>F0 (Hz)</strong></td>
<td>164.80</td>
<td>160.11</td>
<td>113.21</td>
</tr>
<tr>
<td><strong>Log pACC (dB)</strong></td>
<td>−0.8731</td>
<td>−0.7518</td>
<td>−3.1988</td>
</tr>
<tr>
<td><strong>HNR (dB)</strong></td>
<td>23.330</td>
<td>23.330</td>
<td>23.330</td>
</tr>
<tr>
<td><strong>Slope (dB/octave)</strong></td>
<td>−3.904</td>
<td>−4.186</td>
<td>−4.092</td>
</tr>
<tr>
<td><strong>H1–H2 (dB)</strong></td>
<td>−9.610</td>
<td>−11.842</td>
<td>−13.531</td>
</tr>
<tr>
<td><strong>F1 (Hz)</strong></td>
<td>493.85</td>
<td>324.88</td>
<td>375.35</td>
</tr>
<tr>
<td><strong>PE1 (dB)</strong></td>
<td>7.27660</td>
<td>6.50561</td>
<td>3.91152</td>
</tr>
<tr>
<td><strong>F2 (Hz)</strong></td>
<td>2033.27</td>
<td>2383.31</td>
<td>2217.20</td>
</tr>
<tr>
<td><strong>PE2 (dB)</strong></td>
<td>4.00809</td>
<td>3.00739</td>
<td>5.01534</td>
</tr>
</tbody>
</table>

**Table A.4** Duration, glottal source, and formant parameters of tense voice high falling si (name) in Mpi.
<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mpi: man</td>
<td>plain mid level si (a color)</td>
<td>391.9-960.1: 568.2 ms</td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>140.60</td>
<td>140.60</td>
<td>138.58</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>-0.7694</td>
<td>-0.9337</td>
<td>-0.9081</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>23.330</td>
<td>23.330</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>-4.186</td>
<td>-4.657</td>
<td>-4.845</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>-14.666</td>
<td>-16.825</td>
<td>-17.470</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>421.31</td>
<td>281.19</td>
<td>281.19</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>7.03995</td>
<td>6.08241</td>
<td>6.53451</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>1975.38</td>
<td>2092.85</td>
<td>2092.85</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>5.21422</td>
<td>5.10137</td>
<td>6.04580</td>
</tr>
</tbody>
</table>

**Table A.5** Duration, glottal source, and formant parameters of plain voice mid level si (color) in Mpi.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mpi: man</td>
<td>tense mid level si (classifier)</td>
<td>294.0-879.3: 585.3 ms</td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>144.72</td>
<td>140.60</td>
<td>140.60</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>-1.0438</td>
<td>-0.8568</td>
<td>-0.8751</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>23.330</td>
<td>23.330</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>-3.622</td>
<td>-3.998</td>
<td>-4.280</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>-14.130</td>
<td>-18.261</td>
<td>-17.648</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>427.44</td>
<td>281.19</td>
<td>281.19</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>7.30988</td>
<td>6.31544</td>
<td>6.55963</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>2185.50</td>
<td>2249.54</td>
<td>2217.29</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>5.56560</td>
<td>7.00544</td>
<td>6.21809</td>
</tr>
</tbody>
</table>

**Table A.6** Duration, glottal source, and formant parameters of tense voice mid level si (classifier) in Mpi.
<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mpi: man</td>
<td>plain mid rising si ‘roll rope’ 316.0-918.5: 602.5 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>134.64</td>
<td>132.71</td>
<td>148.96</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>–0.9856</td>
<td>–0.9342</td>
<td>–1.7713</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>23.330</td>
<td>23.330</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>–4.186</td>
<td>–4.751</td>
<td>–4.751</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>403.45</td>
<td>265.41</td>
<td>285.28</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>6.83370</td>
<td>6.31855</td>
<td>5.37297</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>1919.15</td>
<td>2123.29</td>
<td>2062.84</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>7.04637</td>
<td>3.51011</td>
<td>4.21539</td>
</tr>
</tbody>
</table>

**Table A.7** Duration, glottal source, and formant parameters of plain voice mid rising si ‘to roll rope’ in Mpi.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mpi: man</td>
<td>tense mid rising si ‘smoke’ 338.0-928.3: 590.3 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>136.59</td>
<td>130.80</td>
<td>144.72</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>–0.9772</td>
<td>–0.9271</td>
<td>–1.0371</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>23.330</td>
<td>23.330</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>–3.810</td>
<td>–4.280</td>
<td>–4.092</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>415.27</td>
<td>391.97</td>
<td>391.97</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>6.97015</td>
<td>6.89288</td>
<td>6.97116</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>2217.29</td>
<td>2217.29</td>
<td>2217.29</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>6.11196</td>
<td>6.06092</td>
<td>6.17675</td>
</tr>
</tbody>
</table>

**Table A.8** Duration, glottal source, and formant parameters of tense voice mid rising si ‘to smoke’ in Mpi.
<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mpi: man</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>134.64</td>
<td>118.23</td>
<td>97.29</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>–1.0646</td>
<td>–1.2898</td>
<td>–2.2734</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>23.330</td>
<td>23.330</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>–4.092</td>
<td>–4.657</td>
<td>–5.033</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>409.32</td>
<td>354.28</td>
<td>293.64</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>6.73503</td>
<td>5.03128</td>
<td>5.13958</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>1919.15</td>
<td>2033.27</td>
<td>2092.85</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>6.44245</td>
<td>5.48399</td>
<td>4.78696</td>
</tr>
</tbody>
</table>

**Table A.9** Duration, glottal source, and formant parameters of plain voice low level si ‘blood’ in Mpi.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mpi: man</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>130.80</td>
<td>114.86</td>
<td>70.81</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>–0.9050</td>
<td>–1.5911</td>
<td>–6.1659</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>12.794</td>
<td>11.289</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>–3.716</td>
<td>–3.622</td>
<td>–3.998</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>523.21</td>
<td>354.28</td>
<td>273.19</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>6.45307</td>
<td>6.40638</td>
<td>2.45229</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>2004.12</td>
<td>2282.26</td>
<td>2282.26</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>2.91168</td>
<td>4.84466</td>
<td>6.75450</td>
</tr>
</tbody>
</table>

**Table A.10** Duration, glottal source, and formant parameters of tense voice low level si ‘seven’ in Mpi.
<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mpi: man</td>
<td>plain low rising si ‘putrid’ 279.3-889.1: 609.8 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>118.23</td>
<td>118.23</td>
<td>162.44</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>–1.1187</td>
<td>–1.1606</td>
<td>–1.3234</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>23.330</td>
<td>23.330</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>–4.186</td>
<td>–4.657</td>
<td>–5.033</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>–13.376</td>
<td>–16.932</td>
<td>–7.792</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>479.79</td>
<td>354.28</td>
<td>320.22</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>6.79778</td>
<td>4.30489</td>
<td>4.70486</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>1919.15</td>
<td>2123.29</td>
<td>2092.85</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>6.03528</td>
<td>6.40760</td>
<td>2.41548</td>
</tr>
</tbody>
</table>

Table A.11 Duration, glottal source, and formant parameters of plain voice low rising si ‘to be putrid’ in Mpi.
<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mpi: man</td>
<td>tense low rising si ‘dried’ 1st subseg.</td>
<td>362.5-663.8: 301.3 ms</td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>128.93</td>
<td>118.23</td>
<td>113.21</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>−1.1787</td>
<td>−1.0124</td>
<td>−1.1340</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>23.330</td>
<td>23.330</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>−3.810</td>
<td>−3.622</td>
<td>−4.092</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>−14.836</td>
<td>−19.073</td>
<td>−18.042</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>523.21</td>
<td>354.28</td>
<td>334.40</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>6.31238</td>
<td>6.47659</td>
<td>6.43025</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>2154.17</td>
<td>2349.14</td>
<td>2249.54</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>2.97097</td>
<td>5.74860</td>
<td>5.40648</td>
</tr>
<tr>
<td></td>
<td>tense low rising si ‘dried’ 2nd subseg. 666.2-916.0: 249.8 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>113.21</td>
<td>134.64</td>
<td>148.96</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>−1.3263</td>
<td>−0.9698</td>
<td>−1.8724</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>23.330</td>
<td>23.330</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>−4.092</td>
<td>−4.186</td>
<td>−4.280</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>−18.314</td>
<td>−15.974</td>
<td>−13.798</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>334.40</td>
<td>265.41</td>
<td>281.19</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>6.43025</td>
<td>5.67260</td>
<td>5.47002</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>2249.54</td>
<td>2185.50</td>
<td>2185.50</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>5.40648</td>
<td>6.02546</td>
<td>5.86022</td>
</tr>
<tr>
<td></td>
<td>tense low rising si ‘dried’ 3rd subseg. 918.5-940.5: 22.0 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>81.81</td>
<td>81.81</td>
<td>61.29</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>−5.0859</td>
<td>−3.9027</td>
<td>−4.1144</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>23.330</td>
<td>23.330</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>−4.751</td>
<td>−4.751</td>
<td>−4.657</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>324.88</td>
<td>324.88</td>
<td>324.88</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>5.06986</td>
<td>4.17260</td>
<td>2.73128</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>2217.29</td>
<td>2217.29</td>
<td>2217.29</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>5.97519</td>
<td>5.81916</td>
<td>5.58208</td>
</tr>
</tbody>
</table>

**Table A.12** Duration, glottal source, and formant parameters of tense voice low rising si ‘to be dried up’ in Mpi.
APPENDIX B

ANALYSIS OF THE GUJARATI SPEECH SAMPLES
<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gujarati: woman</td>
<td>a in paṭ ‘obligation’ 1st subseg.* 172.9-338.5: 165.6 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>289.43</td>
<td>277.16</td>
<td>257.86</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>–0.6331</td>
<td>–0.6486</td>
<td>–0.4698</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>23.330</td>
<td>23.330</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>–3.057</td>
<td>–2.916</td>
<td>–3.057</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>5.125</td>
<td>5.652</td>
<td>5.664</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>1157.73</td>
<td>1108.65</td>
<td>1016.63</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>7.53303</td>
<td>6.16008</td>
<td>3.03953</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>1759.87</td>
<td>1376.78</td>
<td>1299.51</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>2.04974</td>
<td>5.21866</td>
<td>9.10578</td>
</tr>
<tr>
<td></td>
<td>a in paṭ ‘obligation’ 2nd subseg. 341.0-437.4: 96.4 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>261.61</td>
<td>265.41</td>
<td>277.16</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>–0.4533</td>
<td>–0.4783</td>
<td>–0.4589</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>23.330</td>
<td>23.330</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>–3.245</td>
<td>–3.434</td>
<td>–4.186</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>6.415</td>
<td>6.383</td>
<td>4.155</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>1031.42</td>
<td>1061.64</td>
<td>1092.75</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>6.18378</td>
<td>8.15044</td>
<td>8.60123</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>1299.51</td>
<td>1318.41</td>
<td>1396.81</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>8.19399</td>
<td>4.24759</td>
<td>1.08382</td>
</tr>
</tbody>
</table>

Table B.1.1 Duration, glottal source, and formant parameters of pure voice paṭ ‘obligation’ in Gujarati (woman).
<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gujarati: woman</td>
<td>a in paŋ ‘mountain’ 1st subseg.* 257.0-370.7: 113.7 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>269.27</td>
<td>257.86</td>
<td>250.51</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>−0.8180</td>
<td>−0.4908</td>
<td>−0.4960</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>23.330</td>
<td>23.330</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>−3.434</td>
<td>−3.528</td>
<td>−3.528</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>6.372</td>
<td>6.333</td>
<td>6.219</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>1046.42</td>
<td>1016.63</td>
<td>1031.42</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>3.28436</td>
<td>2.79216</td>
<td>6.32016</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>1318.41</td>
<td>1299.51</td>
<td>1299.51</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>8.81381</td>
<td>9.06854</td>
<td>8.08911</td>
</tr>
<tr>
<td></td>
<td>a in paŋ ‘mountain’ 2nd subseg. 373.1-590.6: 217.5 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>257.86</td>
<td>265.41</td>
<td>302.25</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>−0.4753</td>
<td>−0.4709</td>
<td>−0.3785</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>23.330</td>
<td>23.330</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>−3.810</td>
<td>−3.810</td>
<td>−5.221</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>6.002</td>
<td>5.350</td>
<td>−2.187</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>1016.63</td>
<td>1061.64</td>
<td>892.73</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>5.74730</td>
<td>8.20132</td>
<td>4.42326</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>1280.88</td>
<td>1318.41</td>
<td>1208.99</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>7.25507</td>
<td>3.77463</td>
<td>1.42225</td>
</tr>
</tbody>
</table>

*Table B.1.2* Duration, glottal source, and formant parameters of breathy voice paŋ ‘mountain’ in Gujarati (woman).
<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gujarati: woman</td>
<td>a in bar ‘twelve’ 227.3-489.3: 262 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>257.86</td>
<td>261.61</td>
<td>261.61</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>–1.4640</td>
<td>–0.4752</td>
<td>–0.4675</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>23.330</td>
<td>23.330</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>–3.810</td>
<td>–2.916</td>
<td>–3.057</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>4.483</td>
<td>1.674</td>
<td>2.166</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>1016.63</td>
<td>1031.42</td>
<td>1031.42</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>6.16972</td>
<td>2.12556</td>
<td>4.75144</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>1280.88</td>
<td>1299.51</td>
<td>1318.41</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>5.23354</td>
<td>9.11097</td>
<td>9.23777</td>
</tr>
</tbody>
</table>

Table B.1.3 Duration, glottal source, and formant parameters of pure voice bar ‘twelve’ in Gujarati (woman).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gujarati: woman</td>
<td>a in băr ‘outside’ 303.9-603.0: 299.1 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>239.89</td>
<td>257.86</td>
<td>273.19</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>–0.8388</td>
<td>–0.4451</td>
<td>–0.4366</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>23.330</td>
<td>23.330</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>–4.657</td>
<td>–3.716</td>
<td>–5.503</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>6.072</td>
<td>4.001</td>
<td>4.217</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>959.57</td>
<td>1016.63</td>
<td>1092.75</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>6.03068</td>
<td>8.02236</td>
<td>7.69772</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>1437.74</td>
<td>1545.38</td>
<td>1919.15</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>3.41141</td>
<td>4.46594</td>
<td>5.12958</td>
</tr>
</tbody>
</table>

Table B.1.4 Duration, glottal source, and formant parameters of breathy voice băr ‘outside’ in Gujarati (woman).
<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gujarati: woman</td>
<td>a in mal ‘goods, stuff, money’ 274.3-504.1: 229.8 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>257.86</td>
<td>269.27</td>
<td>277.16</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>–0.4570</td>
<td>–0.4692</td>
<td>–0.4609</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>23.330</td>
<td>23.330</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>–3.904</td>
<td>–2.916</td>
<td>–2.916</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>8.522</td>
<td>2.052</td>
<td>2.060</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>987.69</td>
<td>1061.64</td>
<td>1092.75</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>3.71190</td>
<td>4.09347</td>
<td>4.88848</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>1226.57</td>
<td>1357.04</td>
<td>1376.78</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>6.16583</td>
<td>8.81087</td>
<td>8.84163</td>
</tr>
</tbody>
</table>

**Table B.1.5** Duration, glottal source, and formant parameters of pure voice mal ‘goods, stuff, money’ in Gujarati (woman).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gujarati: woman</td>
<td>a in mal ‘rejoice, enjoy’ 341.0-615.4: 274.4 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>239.89</td>
<td>257.86</td>
<td>277.16</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>–0.8388</td>
<td>–0.4679</td>
<td>–0.4254</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>23.330</td>
<td>23.330</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>–4.657</td>
<td>–3.716</td>
<td>–4.468</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>6.072</td>
<td>7.951</td>
<td>7.175</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>1031.42</td>
<td>1031.42</td>
<td>1077.09</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>8.26322</td>
<td>7.80623</td>
<td>7.76378</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>1299.51</td>
<td>1299.51</td>
<td>1947.06</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>3.57887</td>
<td>5.79059</td>
<td>3.03544</td>
</tr>
</tbody>
</table>

**Table B.1.6** Duration, glottal source, and formant parameters of breathy voice mal ‘rejoice, enjoy’ in Gujarati (woman).
<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gujarati: woman</td>
<td>e in pelo ‘that one’ 180.3-316.3: 136.0 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>281.19</td>
<td>273.19</td>
<td>269.27</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>–0.5777</td>
<td>–0.3817</td>
<td>–0.5369</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>23.330</td>
<td>23.330</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>–4.657</td>
<td>–4.751</td>
<td>–5.033</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>–4.373</td>
<td>–6.235</td>
<td>–5.031</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>562.39</td>
<td>546.38</td>
<td>546.38</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>5.64897</td>
<td>7.78772</td>
<td>7.72739</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>2561.75</td>
<td>2753.56</td>
<td>2714.08</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>3.18788</td>
<td>1.82985</td>
<td>1.22924</td>
</tr>
</tbody>
</table>

**Table B.1.7** Duration, glottal source, and formant parameters of pure voice pelo ‘that one’ in Gujarati (woman).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gujarati: woman</td>
<td>će in pelo ‘first’ 185.3-306.4: 121.1 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>285.28</td>
<td>269.27</td>
<td>265.41</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>–0.8034</td>
<td>–0.4584</td>
<td>–0.4429</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>11.289</td>
<td>11.289</td>
<td>11.289</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>–5.691</td>
<td>–6.491</td>
<td>–7.808</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>4.770</td>
<td>4.690</td>
<td>4.580</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>1124.77</td>
<td>1092.75</td>
<td>1061.64</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>5.33773</td>
<td>7.17228</td>
<td>6.09560</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>2282.26</td>
<td>2417.97</td>
<td>2453.14</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>0.81701</td>
<td>4.99399</td>
<td>3.17309</td>
</tr>
</tbody>
</table>

**Table B.1.8** Duration, glottal source, and formant parameters of breathy voice pelo ‘first’ in Gujarati (woman).
Table B.1.9 Duration, glottal source, and formant parameters of pure voice 

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gujarati: woman</td>
<td>e in mel ‘dirt’ 276.7-551.1: 274.4 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>257.86</td>
<td>261.61</td>
<td>265.41</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>–0.6420</td>
<td>–0.4372</td>
<td>–0.4611</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>9.783</td>
<td>12.794</td>
<td>23.330</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>–5.503</td>
<td>–5.033</td>
<td>–6.115</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>2.647</td>
<td>1.655</td>
<td>2.067</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>1016.63</td>
<td>1046.42</td>
<td>1061.64</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>5.98903</td>
<td>6.39704</td>
<td>5.50176</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>2525.03</td>
<td>2561.75</td>
<td>2154.17</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>4.10326</td>
<td>3.49487</td>
<td>2.52846</td>
</tr>
</tbody>
</table>

Table B.1.10 Duration, glottal source, and formant parameters of breathy voice mel ‘palace’ in Gujarati (woman).
<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gujarati: woman</td>
<td>e in wer ‘revenge, animosity’ 291.6-575.8: 284.2 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>257.86</td>
<td>265.41</td>
<td>269.27</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>–0.4366</td>
<td>–0.4356</td>
<td>–0.3791</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>23.330</td>
<td>23.330</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>–5.691</td>
<td>–5.597</td>
<td>–5.974</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>–0.138</td>
<td>0.038</td>
<td>–0.348</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>1262.51</td>
<td>1318.41</td>
<td>1077.09</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>3.10784</td>
<td>4.84066</td>
<td>3.60697</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>2282.26</td>
<td>2383.31</td>
<td>2417.97</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>5.52415</td>
<td>3.07437</td>
<td>2.60622</td>
</tr>
</tbody>
</table>

**Table B.1.11** Duration, glottal source, and formant parameters of pure voice wer ‘revenge, animosity’ in Gujarati (woman).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gujarati: woman</td>
<td>e in wer ‘sawdust’ 496.7-795.8: 299.1 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>250.51</td>
<td>265.41</td>
<td>277.16</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>–0.4745</td>
<td>–0.4193</td>
<td>–0.4074</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>8.278</td>
<td>14.299</td>
<td>23.330</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>–7.620</td>
<td>–6.679</td>
<td>–6.679</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>6.351</td>
<td>3.071</td>
<td>1.670</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>501.03</td>
<td>530.82</td>
<td>554.32</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>5.42538</td>
<td>6.63925</td>
<td>7.41367</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>1759.87</td>
<td>2383.31</td>
<td>2217.29</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>6.60720</td>
<td>2.29896</td>
<td>4.73181</td>
</tr>
</tbody>
</table>

**Table B.1.12** Duration, glottal source, and formant parameters of breathy voice wer ‘sawdust’ in Gujarati (woman).
<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gujarati: man</td>
<td>a in pa<code>obligation</code> 1st subseg. 116.1-237.2: 121.1 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>174.60</td>
<td>148.96</td>
<td>148.96</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>–4.3463</td>
<td>–1.0016</td>
<td>–0.8242</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>23.330</td>
<td>23.330</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>–1.304</td>
<td>4.143</td>
<td>3.478</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>523.21</td>
<td>604.49</td>
<td>587.28</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>2.53810</td>
<td>6.15308</td>
<td>6.35821</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>1031.42</td>
<td>1061.64</td>
<td>1208.99</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>4.49759</td>
<td>2.86201</td>
<td>3.89469</td>
</tr>
<tr>
<td>a in pa<code>obligation</code> 2nd subseg.*</td>
<td>239.7-415.1: 175.4 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>151.12</td>
<td>153.32</td>
<td>174.60</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>–0.9202</td>
<td>–0.8096</td>
<td>–1.2372</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>23.330</td>
<td>23.330</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>–3.245</td>
<td>–3.245</td>
<td>–3.245</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>3.308</td>
<td>3.530</td>
<td>–0.663</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>595.83</td>
<td>613.29</td>
<td>523.21</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>5.91589</td>
<td>6.34722</td>
<td>5.91437</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>1226.57</td>
<td>1244.41</td>
<td>1590.67</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>4.17700</td>
<td>4.48898</td>
<td>4.53122</td>
</tr>
</tbody>
</table>

**Table B.2.1** Duration, glottal source, and formant parameters of pure voice pa`obligation` in Gujarati (man).
<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gujarati: man</td>
<td>a in pə\textsuperscript{d} ‘mountain’ 1st subseg. 202.6-341.0: 138.4 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>177.14</td>
<td>155.55</td>
<td>153.32</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>–1.9181</td>
<td>–0.8631</td>
<td>–0.8209</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>23.330</td>
<td>23.330</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>–3.998</td>
<td>–3.716</td>
<td>–3.622</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>3.329</td>
<td>8.210</td>
<td>8.501</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>718.87</td>
<td>622.21</td>
<td>613.29</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>2.96030</td>
<td>7.19531</td>
<td>7.37175</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>1226.57</td>
<td>1262.51</td>
<td>1226.57</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>1.06392</td>
<td>3.80551</td>
<td>4.03868</td>
</tr>
<tr>
<td></td>
<td>a in pə\textsuperscript{d} ‘mountain’ 2nd subseg.* 343.5-546.1: 202.6 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>153.32</td>
<td>174.60</td>
<td>179.72</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>–0.8205</td>
<td>–1.0819</td>
<td>–0.5903</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>23.330</td>
<td>23.330</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>–3.528</td>
<td>–3.434</td>
<td>–3.810</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>7.790</td>
<td>–0.945</td>
<td>–4.529</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>613.29</td>
<td>530.82</td>
<td>538.54</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>7.66379</td>
<td>6.24061</td>
<td>7.78483</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>1244.41</td>
<td>1637.28</td>
<td>1661.09</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>3.25286</td>
<td>3.71039</td>
<td>3.99448</td>
</tr>
</tbody>
</table>

Table B.2.2 Duration, glottal source, and formant parameters of breathy voice pə\textsuperscript{d} ‘mountain’ in Gujarati (man).
<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gujarati: man</td>
<td>a in bar ‘twelve’</td>
<td>172.9-368.2</td>
<td>195.3 ms</td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>148.96</td>
<td>148.96</td>
<td>142.64</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>-2.0406</td>
<td>-1.4792</td>
<td>-1.2819</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>23.330</td>
<td>23.330</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>-3.810</td>
<td>-3.810</td>
<td>-3.622</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>2.921</td>
<td>2.816</td>
<td>1.175</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>604.49</td>
<td>595.83</td>
<td>562.39</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>5.26487</td>
<td>5.33275</td>
<td>4.95298</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>1046.42</td>
<td>1046.42</td>
<td>1280.88</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>3.82175</td>
<td>3.53201</td>
<td>1.92566</td>
</tr>
</tbody>
</table>

**Table B.2.3** Duration, glottal source, and formant parameters of pure voice bar ‘twelve’ in Gujarati (man).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gujarati: man</td>
<td>a in bar ‘outside’ 1st subseg.</td>
<td>219.9-353.4</td>
<td>213.5 ms</td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>151.12</td>
<td>148.96</td>
<td>144.72</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>-1.5538</td>
<td>-0.8569</td>
<td>-0.8839</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>23.330</td>
<td>23.330</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>-3.716</td>
<td>-3.622</td>
<td>-3.810</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>6.054</td>
<td>6.343</td>
<td>8.509</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>604.49</td>
<td>595.83</td>
<td>578.87</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>5.44036</td>
<td>7.14934</td>
<td>7.60678</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>1226.57</td>
<td>1208.99</td>
<td>1318.41</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>2.55035</td>
<td>3.01114</td>
<td>2.33192</td>
</tr>
</tbody>
</table>

**Table B.2.4** Duration, glottal source, and formant parameters of breathy voice bar ‘outside’ in Gujarati (man).
<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gujarati: man</td>
<td>a in mal ‘goods, stuff, money’ 217.4-467.1: 249.7 ms</td>
<td>146.82</td>
<td>174.60</td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td></td>
<td>174.60</td>
<td></td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>−0.9488</td>
<td>−1.0295</td>
<td>−1.0582</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>23.330</td>
<td>23.330</td>
</tr>
<tr>
<td>Slope (db/octave)</td>
<td>−3.434</td>
<td>−3.151</td>
<td>−3.245</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>3.900</td>
<td>0.823</td>
<td>0.272</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>439.97</td>
<td>698.40</td>
<td>698.40</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>5.66948</td>
<td>5.79922</td>
<td>5.86520</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>1016.63</td>
<td>1417.12</td>
<td>1417.12</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>4.23103</td>
<td>4.55016</td>
<td>4.91642</td>
</tr>
</tbody>
</table>

Table B.2.5 Duration, glottal source, and formant parameters of pure voice mal ‘goods, stuff, money’ in Gujarati (man).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gujarati: man</td>
<td>a in mal ‘rejoice, enjoy’ 232.2-531.3: 299.1 ms</td>
<td>146.82</td>
<td>174.60</td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td></td>
<td>174.60</td>
<td></td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>−0.8665</td>
<td>−0.7254</td>
<td>−0.7631</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>23.330</td>
<td>23.330</td>
</tr>
<tr>
<td>Slope (db/octave)</td>
<td>−4.845</td>
<td>−3.434</td>
<td>−3.622</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>6.845</td>
<td>0.084</td>
<td>−1.636</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>879.93</td>
<td>688.39</td>
<td>698.40</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>4.81581</td>
<td>6.93989</td>
<td>6.74065</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>1479.87</td>
<td>1396.81</td>
<td>1417.12</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>0.23428</td>
<td>5.25995</td>
<td>5.88612</td>
</tr>
</tbody>
</table>

Table B.2.6 Duration, glottal source, and formant parameters of breathy voice mal ‘rejoice, enjoy’ in Gujarati (man).
<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gujarati: man</td>
<td>e in pelo ‘that one’ 111.1-266.8: 155.7 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>169.63</td>
<td>162.44</td>
<td>160.11</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>–3.0130</td>
<td>–1.1322</td>
<td>–0.8615</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>23.330</td>
<td>23.330</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>–4.092</td>
<td>–4.374</td>
<td>–5.033</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>–5.857</td>
<td>–5.116</td>
<td>0.241</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>508.32</td>
<td>493.85</td>
<td>479.79</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>2.46949</td>
<td>4.43131</td>
<td>7.25837</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>2004.12</td>
<td>2062.84</td>
<td>1919.15</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>3.50570</td>
<td>2.84098</td>
<td>1.06686</td>
</tr>
</tbody>
</table>

**Table B.2.7** Duration, glottal source, and formant parameters of pure voice pelo ‘that one’ in Gujarati (man).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gujarati: man</td>
<td>ñ in pelo ‘first’ 222.4-402.8: 180.4 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>179.72</td>
<td>151.12</td>
<td>153.32</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>–1.2207</td>
<td>–0.8487</td>
<td>–0.9014</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>23.330</td>
<td>23.330</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>–4.657</td>
<td>–3.998</td>
<td>–5.597</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>–5.857</td>
<td>–5.116</td>
<td>0.241</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>508.32</td>
<td>493.85</td>
<td>479.79</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>2.46949</td>
<td>4.43131</td>
<td>7.25837</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>2004.12</td>
<td>2062.84</td>
<td>1919.15</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>3.50570</td>
<td>2.84098</td>
<td>1.06686</td>
</tr>
</tbody>
</table>

**Table B.2.8** Duration, glottal source, and formant parameters of breathy voice pelo ‘first’ in Gujarati (man).
**Table B.2.9** Duration, glottal source, and formant parameters of pure voice mel ‘dirt’ in Gujarati (man).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gujarati: man</td>
<td>e in mel ‘dirt’</td>
<td>266.8-548.6: 281.8 ms</td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>148.96</td>
<td>167.20</td>
<td>172.10</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>–0.8879</td>
<td>–0.8887</td>
<td>–0.6678</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>23.330</td>
<td>23.330</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>–3.810</td>
<td>–3.528</td>
<td>–3.810</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>4.441</td>
<td>–0.8887</td>
<td>–1.458</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>446.37</td>
<td>668.79</td>
<td>515.71</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>7.07849</td>
<td>7.23641</td>
<td>7.48252</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>1947.06</td>
<td>1837.78</td>
<td>1709.77</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>4.61294</td>
<td>5.27723</td>
<td>3.69038</td>
</tr>
</tbody>
</table>

**Table B.2.10** Duration, glottal source, and formant parameters of breathy voice mel ‘palace’ in Gujarati (man).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gujarati: man</td>
<td>ê in mël ‘palace’</td>
<td>207.5-449.8: 242.3 ms</td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>144.72</td>
<td>164.80</td>
<td>172.10</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>–0.8849</td>
<td>–1.2441</td>
<td>–0.8250</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>12.794</td>
<td>23.330</td>
<td>23.330</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>–6.867</td>
<td>–3.528</td>
<td>–3.904</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>10.415</td>
<td>4.766</td>
<td>1.243</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>433.66</td>
<td>649.75</td>
<td>515.71</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>5.94319</td>
<td>6.38728</td>
<td>7.03290</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>1975.38</td>
<td>1811.43</td>
<td>1759.87</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>4.03222</td>
<td>4.97482</td>
<td>4.42410</td>
</tr>
<tr>
<td>Parameters</td>
<td>Initial Measure</td>
<td>Mid Measure</td>
<td>Final Measure</td>
</tr>
<tr>
<td>----------------------</td>
<td>-----------------</td>
<td>-------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Gujarati: man</td>
<td>e in wer ‘revenge, animosity’ 224.8-474.5: 249.7 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>142.64</td>
<td>142.64</td>
<td>179.72</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>−0.8594</td>
<td>−0.8069</td>
<td>−0.8046</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>23.330</td>
<td>23.330</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>−3.716</td>
<td>−3.716</td>
<td>−3.904</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>0.422</td>
<td>−0.486</td>
<td>−1.914</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>570.57</td>
<td>570.57</td>
<td>530.82</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>7.46097</td>
<td>8.04412</td>
<td>7.47771</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>1837.78</td>
<td>1837.78</td>
<td>1947.06</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>2.74378</td>
<td>3.91966</td>
<td>3.16073</td>
</tr>
</tbody>
</table>

**Table B.2.11** Duration, glottal source, and formant parameters of pure voice wer ‘revenge, animosity’ in Gujarati (man).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gujarati: man</td>
<td>e in wer ‘sawdust’ 348.4-635.1: 286.7 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>144.72</td>
<td>151.12</td>
<td>177.14</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>−1.0207</td>
<td>−1.0112</td>
<td>−0.7899</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>23.330</td>
<td>23.330</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>−4.280</td>
<td>−3.528</td>
<td>−3.810</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>5.566</td>
<td>5.632</td>
<td>−0.603</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>578.87</td>
<td>604.49</td>
<td>530.82</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>5.73504</td>
<td>7.66881</td>
<td>7.53944</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>1891.63</td>
<td>1947.06</td>
<td>1919.15</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>2.39485</td>
<td>5.77232</td>
<td>3.93293</td>
</tr>
</tbody>
</table>

**Table B.2.12** Duration, glottal source, and formant parameters of breathy voice wer ‘sawdust’ in Gujarati (man).
APPENDIX C

ANALYSIS OF THE JALAPA MAZATEC SPEECH SAMPLES
<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mazatec: woman</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>257.86</td>
<td>254.16</td>
<td>243.38</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>−0.5649</td>
<td>−0.5569</td>
<td>−0.9232</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>23.330</td>
<td>23.330</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>−4.562</td>
<td>−4.468</td>
<td>−6.303</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>−7.224</td>
<td>−7.502</td>
<td>2.653</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>515.71</td>
<td>515.71</td>
<td>750.70</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>5.67120</td>
<td>7.30786</td>
<td>9.07570</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>1545.38</td>
<td>1523.23</td>
<td>1759.87</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>1.15069</td>
<td>3.01928</td>
<td>1.55839</td>
</tr>
</tbody>
</table>

**Table C.1.1** Duration, glottal source, and formant parameters of pure voice sa³ ‘moon’ in Jalapa Mazatec (woman).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mazatec: woman</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>204.66</td>
<td>201.73</td>
<td>201.73</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>−0.8052</td>
<td>−0.5500</td>
<td>−1.2679</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>23.330</td>
<td>23.330</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>−5.315</td>
<td>−4.280</td>
<td>−8.184</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>−0.579</td>
<td>−3.365</td>
<td>7.273</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>613.29</td>
<td>613.29</td>
<td>631.26</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>4.18398</td>
<td>7.56454</td>
<td>9.74303</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>1864.51</td>
<td>2004.12</td>
<td>2123.29</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>0.81366</td>
<td>4.15742</td>
<td>1.06284</td>
</tr>
</tbody>
</table>

**Table C.1.2** Duration, glottal source, and formant parameters of pure voice sæ² ‘he sings’ in Jalapa Mazatec (woman).
<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mazatec: woman</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>254.16</td>
<td>239.89</td>
<td>207.64</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>−1.4521</td>
<td>−1.7640</td>
<td>−3.7929</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>23.330</td>
<td>23.330</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>−4.186</td>
<td>−4.092</td>
<td>−3.528</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>−5.435</td>
<td>−11.257</td>
<td>−7.135</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>523.21</td>
<td>493.85</td>
<td>659.20</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>3.31520</td>
<td>6.68657</td>
<td>3.37599</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>1947.06</td>
<td>1734.64</td>
<td>1734.64</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>1.89739</td>
<td>1.59553</td>
<td>4.19638</td>
</tr>
<tr>
<td><strong>harsh ạ in ʧja³ ‘load’ 1st subseg.</strong> 257.0-350.9: 93.9 ms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>201.73</td>
<td>257.86</td>
<td>265.41</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>−2.7864</td>
<td>−4.1373</td>
<td>−0.7642</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>23.330</td>
<td>23.330</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>−3.434</td>
<td>−3.434</td>
<td>−3.998</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>−7.878</td>
<td>−4.047</td>
<td>−0.983</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>587.28</td>
<td>640.44</td>
<td>515.71</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>5.55491</td>
<td>2.43535</td>
<td>5.42587</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>1734.64</td>
<td>1637.28</td>
<td>1613.80</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>4.52383</td>
<td>4.21374</td>
<td>3.46084</td>
</tr>
<tr>
<td>**harsh ạ in ʧja³ ‘load’ 2nd subseg. 353.4-410.2: 56.8 ms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>261.61</td>
<td>250.51</td>
<td>236.45</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>−0.9613</td>
<td>−1.3123</td>
<td>−1.7321</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>11.289</td>
<td>5.268</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>−5.974</td>
<td>−9.595</td>
<td>−11.477</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>1.763</td>
<td>6.002</td>
<td>13.700</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>530.82</td>
<td>523.21</td>
<td>515.71</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>8.76943</td>
<td>8.41509</td>
<td>4.37025</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>1613.80</td>
<td>1919.15</td>
<td>1919.15</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>3.65998</td>
<td>2.55885</td>
<td>2.74884</td>
</tr>
<tr>
<td>**harsh ạ in ʧja³ ‘load’ 3rd subseg. 412.7-439.9: 27.2 ms</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table C.1.3** Duration, glottal source, and formant parameters of harsh voice ʧja³ ‘load, burden’ in Jalapa Mazatec (woman).
<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mazatec: woman</td>
<td>harsh æ in ñdæ¹ ‘buttocks’ 286.6-400.3: 113.7 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>111.59</td>
<td>177.14</td>
<td>184.98</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>–1.6851</td>
<td>–2.8234</td>
<td>–1.6226</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>23.330</td>
<td>11.289</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>–3.528</td>
<td>–3.716</td>
<td>–12.229</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>–7.028</td>
<td>–1.248</td>
<td>8.644</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>678.52</td>
<td>708.56</td>
<td>905.72</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>6.39370</td>
<td>4.60677</td>
<td>1.72362</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>1891.63</td>
<td>1919.15</td>
<td>2062.84</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>5.20440</td>
<td>4.02216</td>
<td>2.20724</td>
</tr>
</tbody>
</table>

**Table C.1.4** Duration, glottal source, and formant parameters of harsh voice ñdæ¹ ‘buttocks’ in Jalapa Mazatec (woman).
<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mazatec: woman breathy a in &quot;dá 23 'hard' 1st subseg. 338.5-422.6: 84.1 ms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>226.43</td>
<td>239.89</td>
<td>246.92</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>−0.6052</td>
<td>−0.5573</td>
<td>−0.4897</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>15.804</td>
<td>23.330</td>
<td>23.330</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>−6.303</td>
<td>−3.904</td>
<td>−3.998</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>6.577</td>
<td>−1.633</td>
<td>−2.166</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>678.53</td>
<td>718.87</td>
<td>987.69</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>4.53644</td>
<td>6.58248</td>
<td>7.31099</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>1141.13</td>
<td>1437.74</td>
<td>1458.65</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>2.76658</td>
<td>2.63473</td>
<td>3.08792</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mazatec: woman breathy a in &quot;dá 23 'hard' 2nd subseg. 425.0-472.0: 47.0 ms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>246.92</td>
<td>239.89</td>
<td>233.06</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>−0.5111</td>
<td>−0.8560</td>
<td>−2.5085</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>23.330</td>
<td>11.289</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>−4.751</td>
<td>−9.595</td>
<td>−17.874</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>−0.096</td>
<td>3.595</td>
<td>12.533</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>739.93</td>
<td>493.85</td>
<td>486.77</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>8.00683</td>
<td>8.36965</td>
<td>5.81027</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>1479.87</td>
<td>1837.78</td>
<td>1919.15</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>2.22395</td>
<td>2.83003</td>
<td>3.14641</td>
</tr>
</tbody>
</table>

**Table C.1.5** Duration, glottal source, and formant parameters of breathy voice "dá 23 'hard' in Jalapa Mazatec (woman).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mazatec: woman breathy æ in &quot;dá 1 'horse' 254.5-380.5: 126.0 ms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>195.98</td>
<td>193.17</td>
<td>167.20</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>−0.7964</td>
<td>−0.6865</td>
<td>−1.5767</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>23.330</td>
<td>9.783</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>−6.491</td>
<td>−5.315</td>
<td>−16.369</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>4.282</td>
<td>1.152</td>
<td>11.794</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>587.28</td>
<td>587.28</td>
<td>570.57</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>3.60473</td>
<td>4.96792</td>
<td>1.79420</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>2062.84</td>
<td>2123.29</td>
<td>2123.29</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>2.16003</td>
<td>2.67740</td>
<td>4.27136</td>
</tr>
</tbody>
</table>

**Table C.1.6** Duration, glottal source, and formant parameters of breathy voice "dá 1 'horse' in Jalapa Mazatec (woman).
<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mazatec: man</td>
<td>pure a in sa³ ‘moon’ 232.2-412.7: 180.5 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>229.72</td>
<td>226.43</td>
<td>207.64</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>–0.7804</td>
<td>–0.4836</td>
<td>–1.2271</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>23.330</td>
<td>17.309</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>–4.280</td>
<td>–3.810</td>
<td>–17.874</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>–3.808</td>
<td>–2.255</td>
<td>15.306</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>688.39</td>
<td>678.52</td>
<td>668.79</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>5.45226</td>
<td>7.99641</td>
<td>5.27427</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>1376.78</td>
<td>1590.67</td>
<td>1837.78</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>2.21597</td>
<td>3.43681</td>
<td>0.90701</td>
</tr>
</tbody>
</table>

**Table C.2.1** Duration, glottal source, and formant parameters of pure voice sa³ ‘moon’ in Jalapa Mazatec (man).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mazatec: man</td>
<td>pure æ in sæ² ‘sings’ 1st subseg.* 257.0-452.2: 195.2 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>153.32</td>
<td>157.81</td>
<td>160.11</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>–1.1018</td>
<td>–0.8737</td>
<td>–0.8000</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>23.330</td>
<td>23.330</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>–4.562</td>
<td>–3.810</td>
<td>–4.092</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>–0.349</td>
<td>1.472</td>
<td>1.336</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>459.45</td>
<td>795.33</td>
<td>806.90</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>4.47733</td>
<td>6.00955</td>
<td>7.06750</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>1785.46</td>
<td>1891.63</td>
<td>1919.15</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>3.61344</td>
<td>5.27560</td>
<td>6.24720</td>
</tr>
</tbody>
</table>

**Table C.2.2** Duration, glottal source, and formant parameters of pure voice sæ² ‘he sings’ in Jalapa Mazatec (man).
<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mazatec: man</td>
<td>harsh a in tʃa³ ‘load, burden’ 219.9-343.5: 123.6 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>223.18</td>
<td>207.64</td>
<td>198.83</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>–3.3625</td>
<td>–0.7160</td>
<td>–2.4893</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>23.330</td>
<td>8.278</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>–3.434</td>
<td>–3.434</td>
<td>–11.853</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>–2.794</td>
<td>–0.099</td>
<td>11.567</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>678.52</td>
<td>818.64</td>
<td>1031.42</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>6.98238</td>
<td>3.86260</td>
<td>1.91731</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>1637.28</td>
<td>1226.57</td>
<td>1685.25</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>4.62608</td>
<td>4.04622</td>
<td>3.68101</td>
</tr>
</tbody>
</table>

Table C.2.3 Duration, glottal source, and formant parameters of harsh voice tʃa³ ‘load, burden’ in Jalapa Mazatec (man).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mazatec: man</td>
<td>harsh æ in nɒæ⁴ ‘buttocks’ 333.6-449.8: 116.2 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>146.82</td>
<td>148.96</td>
<td>127.08</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>–0.9667</td>
<td>–1.3810</td>
<td>–4.0060</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>23.330</td>
<td>23.330</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>–3.810</td>
<td>–3.716</td>
<td>–4.186</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>–5.076</td>
<td>–1.292</td>
<td>2.238</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>739.93</td>
<td>750.70</td>
<td>729.32</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>7.25964</td>
<td>8.73890</td>
<td>4.01950</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>1785.46</td>
<td>1919.15</td>
<td>1919.15</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>4.36265</td>
<td>2.24630</td>
<td>4.29821</td>
</tr>
</tbody>
</table>

Table C.2.4 Duration, glottal source, and formant parameters of harsh voice nɒæ⁴ ‘buttocks’ in Jalapa Mazatec (man).
<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mazatec: man</td>
<td>breathy ( \ddag ) in (^\ddag )ˈdà(^{23}) ‘hard’ 1st subseg.</td>
<td>294.0-427.5: 133.5 ms</td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>125.26</td>
<td>195.98</td>
<td>201.73</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>−1.5209</td>
<td>−0.8326</td>
<td>−0.7369</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>9.783</td>
<td>23.330</td>
<td>18.814</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>−8.372</td>
<td>−5.409</td>
<td>−5.974</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>5.066</td>
<td>4.672</td>
<td>4.873</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>1002.06</td>
<td>772.69</td>
<td>783.93</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>3.82674</td>
<td>4.12775</td>
<td>4.82700</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>1685.25</td>
<td>1637.28</td>
<td>1637.28</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>4.77493</td>
<td>2.73462</td>
<td>3.07258</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>breathy ( \ddag ) in (^\ddag )ˈdà(^{23}) ‘hard’ 2nd subseg.</td>
<td>430.0-486.8: 56.8 ms</td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>201.73</td>
<td>190.40</td>
<td>177.14</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>−0.7595</td>
<td>−1.1752</td>
<td>−2.8136</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>20.320</td>
<td>23.330</td>
<td>12.794</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>−12.229</td>
<td>−17.874</td>
<td>−13.358</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>8.290</td>
<td>12.576</td>
<td>9.860</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>1016.63</td>
<td>1031.42</td>
<td>1046.42</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>5.88711</td>
<td>2.14310</td>
<td>2.01976</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>1637.28</td>
<td>1458.65</td>
<td>1479.87</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>2.19182</td>
<td>1.85934</td>
<td>3.07792</td>
</tr>
</tbody>
</table>

Table C.2.5 Duration, glottal source, and formant parameters of breathy voice \(^\ddag \)ˈdà\(^{23}\) ‘hard’ in Jalapa Mazatec (man).
<table>
<thead>
<tr>
<th>Parameters</th>
<th>Initial Measure</th>
<th>Mid Measure</th>
<th>Final Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mazatec: man</td>
<td>breathy æ in &quot;dæ&quot; 'horse' 1st subseg.* 224.8-353.4: 128.6 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>89.21</td>
<td>116.53</td>
<td>118.23</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>–2.3871</td>
<td>–1.1307</td>
<td>–1.1300</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>23.330</td>
<td>18.814</td>
<td>15.804</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>–2.916</td>
<td>–5.691</td>
<td>–5.974</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>1.376</td>
<td>2.807</td>
<td>3.525</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>698.40</td>
<td>688.39</td>
<td>698.40</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>3.89853</td>
<td>4.04762</td>
<td>4.31205</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>1947.06</td>
<td>1919.15</td>
<td>1891.63</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>6.13013</td>
<td>4.13578</td>
<td>4.53750</td>
</tr>
<tr>
<td></td>
<td>breathy æ in &quot;dæ&quot; 'horse' 2nd subseg. 355.8-422.6: 66.8 ms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F0 (Hz)</td>
<td>114.86</td>
<td>106.86</td>
<td>100.14</td>
</tr>
<tr>
<td>Log pACC (dB)</td>
<td>–1.3161</td>
<td>–1.7873</td>
<td>–7.7167</td>
</tr>
<tr>
<td>HNR (dB)</td>
<td>14.299</td>
<td>17.309</td>
<td>18.814</td>
</tr>
<tr>
<td>Slope (dB/octave)</td>
<td>–7.996</td>
<td>–11.477</td>
<td>–17.874</td>
</tr>
<tr>
<td>H1–H2 (dB)</td>
<td>5.734</td>
<td>7.496</td>
<td>7.824</td>
</tr>
<tr>
<td>F1 (Hz)</td>
<td>698.40</td>
<td>678.52</td>
<td>783.93</td>
</tr>
<tr>
<td>PE1 (dB)</td>
<td>5.18742</td>
<td>4.34439</td>
<td>2.68181</td>
</tr>
<tr>
<td>F2 (Hz)</td>
<td>1891.63</td>
<td>1919.15</td>
<td>1919.15</td>
</tr>
<tr>
<td>PE2 (dB)</td>
<td>4.66251</td>
<td>4.52226</td>
<td>8.04580</td>
</tr>
</tbody>
</table>

Table C.2.6 Duration, glottal source, and formant parameters of breathy voice "dæ" 'horse' in Jalapa Mazatec (man).
REFERENCES


VITA

1985…………….Bachelor of Arts in Linguistics (Licence), University of Provence I

1987…………….Masters of Arts in Phonetics (D.E.A.), University of Provence I

PUBLICATIONS


Travaux de l’Institut de Phonétique d’Aix 11, 115-154.