The timing of nonmodal phonation in vowels

Barbara Blankenship

Phonetics Laboratory, Linguistics Department, UCLA, Los Angeles, CA 90024-1543, U.S.A.

In a language where breathiness or laryngealization is a contrastive property of vowels, such nonmodal phonation lasts longer and is more differentiated from modal phonation than in a language where nonmodal phonation results from the influence of preceding consonants. In Tagalog, breathy phonation occurs incidentally on vowels after /h/, and laryngealized phonation occurs after glottal stops. Mazatec, on the other hand, employs breathy and laryngealized vowels as separate phonemes that contrast with modal vowels. Several acoustic measures show that nonmodal and modal vowels are differentiated more strongly and over a longer duration in Mazatec than in Tagalog.

An experiment examined words from six male and six female speakers of each of those languages, with corroborating modal and breathy vowels from four male and four female speakers of Chong, and modal and laryngealized vowels from one male speaker of Mpi. From each speaker, three vowels of each phonation type were analyzed. To determine the time course of phonation effects, measurements were made at 25 ms intervals through each vowel. The measurements were the amplitude differences between the first and second harmonic and between the first harmonic and the second formant, and cepstral peak prominence (a measure of periodicity).

© 2002 Elsevier Science Ltd. All rights reserved.

1. Introduction

Studies of the linguistic uses of voice source variation have dealt in some detail with the time course of fundamental frequency (f_0) both as it relates to tone in tone languages and to phrasal pitch patterns in nontone languages. F_0 is controlled by the *rate* of vibration of the vocal folds. Less attention has been given to the linguistic use of variations in the *manner* of vibration: modal (the standard vibration type), breathy (where the folds are held apart so that the glottis fully closes for only a very small portion of the vibration cycle, if at all), and laryngealized (where the folds are held stiffly and vibration is partially inhibited.) These manners of vibration are referred to collectively as phonation types. Ladefoged (1983) gives an introduction to the linguistic uses of phonation type. Excellent photographs of a glottis producing the three types of phonation can be seen in Ladefoged (2001).

Use of the term "laryngealized" requires some explanation. Laryngealization often results in an audible creaky sound, but since that is not always the case, I have elected not to use the term creaky. Laryngealization can cause the arytenoid and ligamental portions of the vocal folds to vibrate out of phase with each other, producing pulses with alternating high and low amplitudes that are perceived as a creaking sound (Ladefoged & Maddieson, 1996). In Mazatec and Mpi, laryngealized vowels do not consistently have an audible creak nor display irregular glottal pulses on a spectrogram. Creakiness appears to be an occasional side effect of the laryngealization rather than its goal. Fig. 1(a) shows the spectrogram and spectrum of a *noncreaky* laryngealized vowel. Fig. 1(b) shows a *creaky* laryngealized vowel. Although the spectrograms of the two vowels
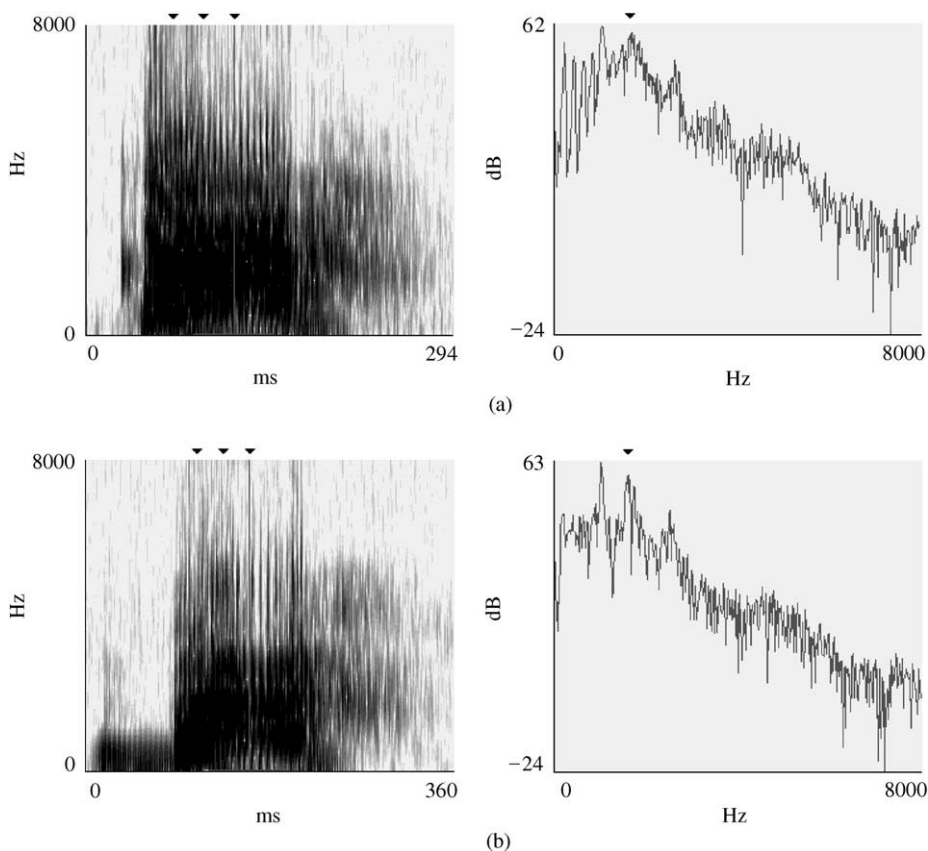

Figure 1. Four Mazatec monosyllables, from recordings of speaker 3, female. For each word, the left window shows a spectrogram of the waveform, and the right window shows an FFT spectrum of the waveform, taken over a 25 ms interval centered at the 75th ms of the vowel (indicated by arrows above the spectrogram). Arrows above the spectra point to F2. (a) Laryngealized vowel in the word [ⁿka̰ ²] "high"; (b) laryngealized vowel in the word [βa̰²] "carries"; (c) breathy vowel in the word [ⁿdja̤: ²³] "cornflour drink"; (d) modal vowel in the word [ka ²] "bald". Note that the laryngealized vowel in Fig. 1(a) does not exhibit creakiness in the spectrogram, but it does have a spectrum similar to that of the other laryngealized vowel in Fig. 1(b): the difference $H_1 - H_2$ is considerably less in the laryngealized vowels than in the modal vowel in Fig. 1(d); the difference $H_1 - F_2$ is negative in the laryngealized vowels, positive in the modal vowel.
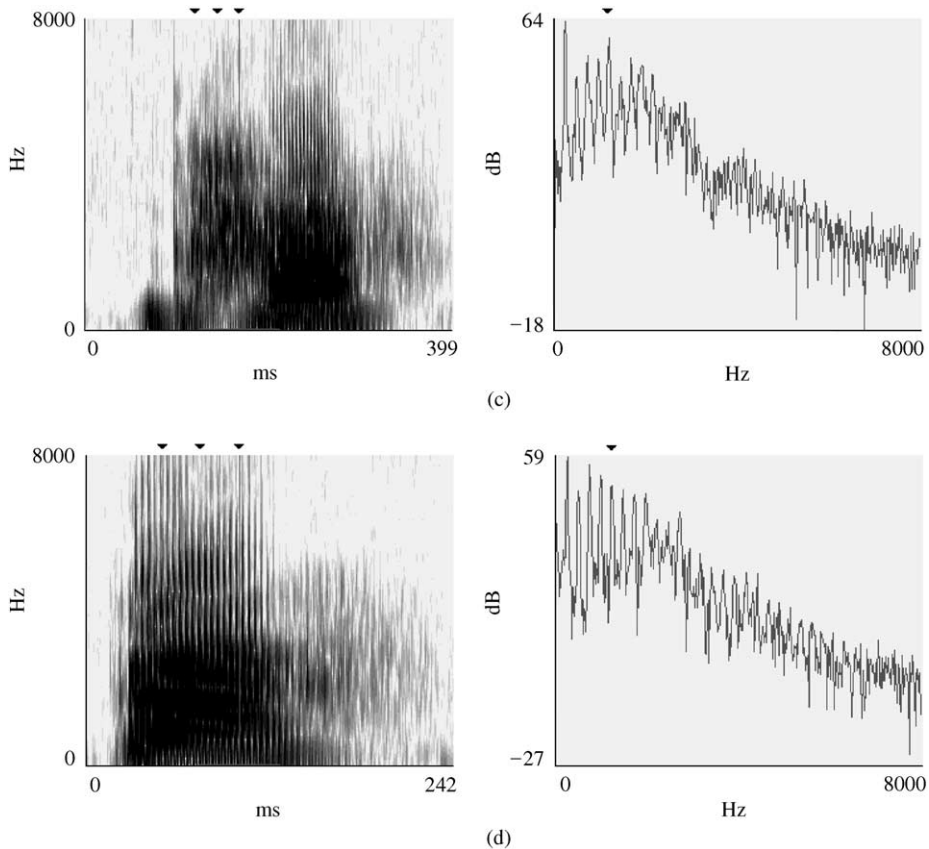
Figure 1 (Continued).

are different, the spectra are similar. It is the spectral shape that distinguishes the vowel's phonation type in Mazatec.

In this study, all laryngealized vowels were considered as a single phonation type, whether they were creaky or not. Titze (1995) and Stevens (1999) use the term "pressed voice" for this type of phonation. Pierrehumbert & Talkin (1992, p. 93) refer to a "braced configuration" of the vocal folds. See Laver (1980, 1994), and Ní Chasaide & Gobl (1997) for views on the distinction between laryngealized and creaky phonation.

The time course of variations in phonation type has recently received some attention. Silverman, Blankenship, Kirk & Ladefoged (1995) found that in Mazatec, which has contrastive modal, breathy, and laryngealized vowels, contrastive breathiness lasts for only 43% of the vowel duration, giving way to modal vibration thereafter. It is noteworthy that nonmodal phonation lasts for less than half of the vowel where it is a contrastive feature. By comparison, the feature nasality has been shown (Cohn, 1990) to persist throughout the vowel, both when it is an underlying feature (French) and when it is specified on a vowel by phonological rule (English nasal consonant deletion, Sundanese nasal spreading). Silverman (1995) hypothesized that the long modal portion of the breathy vowel serves to make pitch information more salient.

Languages that do not have phonation specification on vowels can have breathy vowels near [h] or aspirated consonants, and laryngealized vowels near [ʔ] or

laryngealized consonants. For convenience I shall refer to such vowels as "noncontrastive", although clearly the preceding consonant bears a contrast. What do we know about the duration of nonmodal phonation in vowels where phonation type is *not* contrastive? It was found by Löfqvist & McGowan (1992) that vowels after [h] or aspirated [pʰ] have breathy phonation during approximately the initial seven cycles of the vowel for one male speaker of Swedish (120 Hz fundamental frequency) and the initial 11 cycles for one female speaker of English (210 Hz fundamental frequency), corresponding to absolute durations of about 58 and 52 ms, respectively. If for comparison we assume the same average fundamental frequencies (120 and 210 Hz) for Silverman's male and female subjects, Mazatec breathy vowels would have breathy phonation during the initial 13 cycles for males and 23 cycles for females, roughly twice as long as in Löfqvist and McGowan's noncontrastive vowels. (Since the Swedish and English vowel durations were not given, we cannot determine which percentage of the vowel was affected by the phonation change.) The longer breathiness in Mazatec could indicate that Mazatec speakers control phonation duration to make the necessary contrast, or it may be due to differences in methods of the two studies. Löfqvist and McGowan measured the open quotient of the airflow wave of nonsense syllables; Silverman observed narrow-band spectrograms of actual words.

Is the time course of the effects of breathy voice similar across languages that contrast phonation types on vowels? Do other languages that have contrastive breathiness limit the acoustic cues to breathiness to the first half of the vowel in the same way that Mazatec does?

A central thesis of Silverman (1995) is that simultaneous phonological features that would tend to obscure each other in perception may be realized nonsimultaneously in order to maximize the salience of each one. Thus, the Mazatec breathy vowel is breathy initially but becomes modal during the second half in order to render f_0 more perceptible for distinguishing tone. The periodic signal is weak in breathy phonation, making it more difficult for listeners to determine pitch. Since laryngealized phonation has a strong periodic component, tone perception of laryngealized vowels should be less of a problem. Thus, one can predict that Mazatec laryngealized vowels would not become modal during their second half, if perceptual salience is the ruling factor. Results on the time course of laryngealization in Mazatec, therefore, have a bearing on Silverman's theory.

This study looks at how phonemic specification of a phonation type on a vowel affects the time course of that phonation during the vowel. Both breathy and laryngealized phonation are considered, in languages where the phonation is specified on the vowel itself and in languages where nonmodal phonation occurs on vowels as a result of contrasts in the preceding consonant. The study focuses on two questions. No predictions or hypotheses are implied. The questions provide a framework for observations only.

1. Is nonmodal phonation of *longer duration* in languages with contrastive phonation types?
2. Is nonmodal phonation *more different* from modal phonation in languages with contrasting vowel phonation types?

2. Choice of analysis method

2.1. *Background*

A given phonation type may manifest itself in different ways in the acoustic signal because several different glottal actions can be used to achieve the perceptual effect of

TABLE I. Elements of breathy articulation

Glottal state	Postulated effect on glottal vibration	Postulated acoustic outcome
Vocal folds less tense	Larger ratio of open phase to complete cycle (open 80–100% of cycle, *vs.* 65–70% for modal phonation (Childers & Lee, 1991))	Spectrum dominated by F_0, thus H_1 has markedly higher amplitude than the other harmonics (large $H_1 - H_2$ difference)
	Glottal vibration approaches closed phase less abruptly	Steeper spectral slope above 2000 Hz
	Glottal vibration has little or no closed phase	Steeper spectral slope above 2000 Hz; increased F_1 bandwidth
Arytenoids held open, allowing airflow between the arytenoids even when the main portion of the vocal folds is closed	Wave includes both periodic and aperiodic elements	Nonperiodic aspiration noise at high frequencies (can interfere with steeper spectral slope measurement); increased F_1 bandwidth

TABLE II. Possible elements of laryngealized articulation

Glottal state	Possible effect on glottal vibration	Possible acoustic outcome
Vocal folds more tense	Smaller ratio of open phase to complete cycle (open 25–45% of cycle (Childers & Lee, 1991))	H_1 amplitude often less than that of H_2
	Glottal vibration approaches closed phase abruptly, exciting frequencies throughout the spectrum.	Spectral slope above 2000 Hz is more gradual, not as steep
	Glottal vibration has unequal cycles	Jitter

breathiness or laryngealization. A speaker may use one or more of these actions for each phonation type. Table I summarizes some of the glottal states suggested for breathy vowels in Stevens (1999), Stevens & Hanson (1995), Klatt & Klatt (1990), Gobl & Ní Chasaide (1988). The left column lists articulatory configurations, the middle column gives the postulated effect of each configuration on the glottal vibration, and the right column lists the acoustic outcome claimed to result from the articulation. Table II presents a possible glottal state for laryngealized vowels. The items are speculative, but logically related to the theories summarized in Table I.

There is still not a great deal of experimental evidence for the presumed correlation between the glottal effects listed in the middle column of the tables and the acoustic outcomes listed in the right columns. One goal of the study reported in Holmberg Hillman, Perkell, Guiod & Goldman (1995) was to seek correlations between airflow, electroglottographic, and acoustic data, to determine if acoustic data could substitute for

the other kinds of measures in clinical use. The study determined that the amplitude difference between the first two harmonics in the acoustic signal ($H_1 - H_2$) correlates with the open quotient, the percentage of a glottal vibration cycle during which the glottis is open. The open quotient measurement was taken from the airflow waveform, which reflects the actual pattern of glottal opening and closure.

The study of American English gender differences in breathiness by Hanson (1997), found that $H_1 - H_2$ of the acoustic signal did not correlate with more global measures of spectral slope ($H_1 - F_1$ and $H_1 - F_3$). Since $H_1 - H_2$ had been shown by Holmberg *et al.* (1995) to relate to the open quotient, it may be that the more global measure is related to some other glottal parameter. A theory that originated with Stevens (1977) maintains that the slope of the source spectrum correlates with the abruptness or gradualness of vocal fold closure. When the vocal folds come together gradually over their length, they may excite primarily the lower frequencies of the vocal tract, resulting in a steeply sloped spectrum, with most of the energy near f_0 and very little energy at higher frequencies. When the folds come together all at once, they may provide efficient excitation of a wider range of frequencies, producing a spectrum that is less steep, with higher frequency components relatively stronger. While there are few articulatory observations yet to support the theory, it does provide a convenient framework for discussion. In any case, it is prudent to include both types of measurements in a phonation study. For convenience, I shall use the term "spectral slope" to refer to the more global spectral measures but not to $H_1 - H_2$.

Regardless of the articulatory mechanism of spectral slope adjustment, it is a fact that each phonation type has a distinct slope. Slope differences are illustrated in Fig. 1, which shows the FFT spectrum at the 75th ms for Mazatec breathy, modal, and laryngealized [a] vowels spoken by the same speaker. Observe the second formant (marked with an arrow above each spectrum), whose amplitude is greater than that of H_1 in the laryngealized vowel and less than that of H_1 in the breathy and modal vowels.

In actual breathy speech, the speaker may produce a spectrum with a large $H_1 - H_2$ difference, steep slope, aspiration noise, or a combination of those elements. Since we cannot anticipate which cues a speaker will produce, it is desirable to have measures of several possible acoustic results in order to be sure of capturing the contrast.

2.2. *Acoustic measurements*

The data used in this study impose two constraints on the choice of analysis method. First, only tape recordings were available; therefore, only acoustic measures could be used. Second, since the experimental design required comparisons over very short windows (25 ms), methods that require a longer sample were ruled out. In addition, the large number of samples analyzed for each token made it impractical to select measures that would require a great deal of manipulation.

A pilot study (Blankenship, 1997) compared 10 measures on their ability to differentiate among 25 ms samples of breathy, modal, and laryngealized vowels in Mazatec. From the most successful measures, I selected both $H_1 - H_2$ and $H_1 - F_2$, since they may indicate different glottal postures, and cepstral peak amplitude as the indicator of periodicity. I did not select a jitter measurement, since jitter did not differentiate the Mazatec phonation types well in the pilot study.

A cepstrum is a second-order spectrum generated by taking the fast fourier transform (FFT) of the log magnitude values of a power spectrum. The dimensions of a cepstrum

are quefrency (in seconds) and gamnitude (in dB). The spectrum of a periodic signal shows well-defined harmonics; its cepstrum has a prominent peak at a quefrency corresponding to the duration of the f_0 cycle. Less periodic signals such as those often produced in breathy phonation have a spectrum with less definite harmonics, resulting in a cepstrum with a low peak. If modal phonation is more periodic than either breathy or laryngealized phonation, one would expect higher cepstral peaks for modal vowels. The measure is unreliable where there are rapid pitch changes, however, and where the vocal folds of a modal vowel happen to be vibrating irregularly.

Fig. 2(a) shows a spectrum on the left and a cepstrum on the right, calculated over a 25 ms window centered on the 25th ms of the breathy vowel in Fig. 1(c). Fig. 2(b) shows the same measurements for the modal vowel in Fig. 1(d). The spectrum of the breathy vowel is not very periodic. Its cepstrum has a barely discernable peak at 5.2 ms, the cycle duration of a 192 Hz fundamental. The spectrum of the modal vowel has regular peaks at
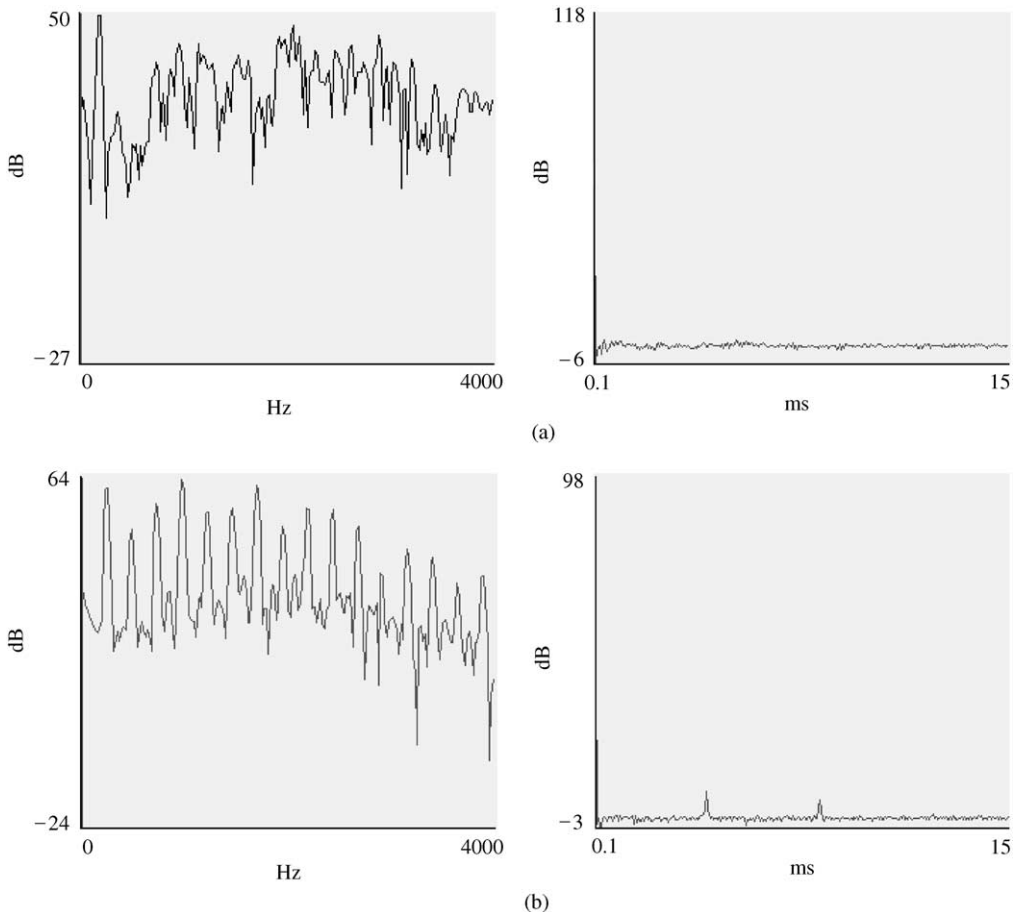

Figure 2. Fast Fourier transform (FFT) spectrum and cepstral representation of breathy and modal vowels in Fig. 1, taken over a 25 ms interval centered at the 25th ms of the vowel. The location of the highest peak on the *x*-axis scale (the "quefrency") corresponds to the fundamental period of the signal: (a) breathy vowel in the word [ⁿdjạː [23]] "cornflour drink"; (b) modal vowel in the word [ka [2]] "bald".

each harmonic. Its cepstrum has a pronounced peak at 4.1 ms, the cycle duration of a 244 Hz fundamental. (The peak at 8.2 ms corresponds to the subharmonics seen between the main harmonics in the spectrum.) The modal peak is 7.87 dB and the breathy peak is 1.86 dB. To normalize for differences in energy, the average amplitude of all the cepstral points was subtracted from the amplitude of the peak; the amplitude difference was used as the unit for comparison.

3. Language materials

The study required one language with contrastive breathiness or laryngealization on vowels, and another language that has only modal vowels but offers consonant environments that induce breathiness or laryngealization on following vowels. For convenience, I shall refer to the two kinds of languages as "contrastive" and "noncontrastive" languages.

The contrastive language was Jalapa Mazatec. The noncontrastive language was Tagalog, which has syllable-initial [ʔ] and [h] in its consonant inventory. From each language I selected three words as exemplars of each phonation type. The target vowel in all words was [a] with lexical stress.

One could also have included another category of noncontrastive languages, those with aspirated and laryngealized obstruents. See Cho, Jun & Ladefoged (2002) for an analysis of such consonants in Korean. The category was omitted from this study in order to keep the analysis as clear as possible. Aspirated and laryngealized consonants require movement of upper vocal tract articulators in addition to the glottis, which introduces confounding effects on the acoustics of the following vowel.

The Mazatec results are more interesting if they can be corroborated with data from other languages. The languages selected for corroboration are unrelated to those used in the main study and are sufficiently distant geographically to rule out similarities due to borrowing or areal trends. A language comparable to Mazatec in having contrastive breathy and modal vowels is Chong. Although Chong also has contrastive laryngealization, the laryngealization occurs near the end of the vowel. Since this study concentrates on vowel onsets, I did not include Chong laryngealized vowels.

Mpi provides another example of contrasting laryngealized and modal vowels. Laryngealized vowels in Mpi do not sound like those of Mazatec, because they have an additional characteristic—perhaps faucal tension or some movement of the tongue root—that is not present in Mazatec. The exact nature of the characteristic is outside the scope of this study.

As far as possible the target vowels were in stressed mid-tone syllables, maximally distant from other consonants that could hold the vocal folds apart (e.g., voiceless consonants) or have acoustic effects on the vowel that would confound the spectral analysis (e.g., nasals, with their effect on $F1$ bandwidth.) Six male and six female adult speakers of Mazatec and of Tagalog provided the main data set. Four male and four female speakers of Chong, and one male speaker of Mpi provide corroborating data.

3.1. *Jalapa Mazatec*

Mazatec belongs to the Popolocan branch of the Otomanguean language family. Jalapa Mazatec is spoken in the vicinity of San Felipe Jalapa de Diaz, in the northeastern foothills on the Gulf side of Oaxaca, Mexico. A 1990 census indicates that there are

TABLE III. The Mazatec sample words. Tone 1 is low, 2 is mid, and 3 is high

Laryngealized		Modal		Breathy	
βa̠^{2}	"carries"	^{n}da^2	"good"	^{n}da̤ː^{23}	"hard"
tʃa̠^{3}	"load, burden"	na^1	"woman"	^{n}dja̤ː^{23}	"cornflour drink"
^{n}ka̰^{2}	"high, tall"	ka^2	"bald"	kiŋa̤ː^{23}	"he fastened"

TABLE IV. The Tagalog sample words. Target vowels are underlined

After [ʔ]		After [b/d/g]		After [h]	
paʔa	"foot"	baga	"embers"	pahajag	"article"
tagaʔajos	"manager"	dagat	"sea"	mahal	"expensive"
paʔalam	"good-bye"	gaga	"fool"(fem.)	mahalai	"depraved"

15 500 speakers, 4600 of them monolingual (Grimes, 1996). The language has a five-vowel system, but tonal, laryngeal, nasal, and length contrasts greatly expand the vowel inventory. Contrastive breathy, modal, and laryngealized phonation occurs on all five vowels and all three tones. Thus, laryngealization is not associated solely with low pitch as in many other languages. A detailed description of the phonetics of Jalapa Mazatec was given by Silverman *et al.* (1995).

Table III gives the Mazatec sample words, which were recorded without a framing sentence. Tone 1 is low, tone 2 mid, and tone 3 high. Since the word list was developed for other purposes, the sample words are not ideally matched for this study, but they are the best set available within the recorded list. Although Mazatec allows both long and short vowels in all phonation types, the breathy vowels that best fit the tone and context criteria for this experiment happened to be long, and the modal and laryngealized vowels short. Paul Kirk and Peter Ladefoged made the recordings in Jalapa de Diaz in April 1993. Most of the male speakers were bilingual in Spanish and Mazatec; most of the females were monolingual.

3.2. *Tagalog*

Tagalog is a member of the western Malayo-Polynesian branch of the Austronesian language family. Originally spoken in the southern part of Luzon, it has spread throughout the Philippines since 1937, when it was selected as the national language (Schachter, 1987). 14 850 000 people speak Tagalog as a first language (Grimes, 1996). Tagalog has five modal vowels; lexical stress is marked primarily by increased vowel duration. The consonants include both [h] and [ʔ]. Although the Tagalog alphabet has no symbol for it, speakers pronounce all orthographically adjacent vowels with an intervening [ʔ]. Unlike English, Tagalog [ʔ] is not in free variation with zero onset in this context. Schachter (1987) contains a description of the Tagalog phonetic inventory.

Table IV shows the Tagalog sample words. Underlines indicate the target vowels, all on stressed syllables. The sample was not controlled for the position of the target syllable within the word. The words were recorded at the UCLA phonetics laboratory during the summer of 1995. All 12 readers were fluent in both English and Tagalog. Two were fluent in Spanish and two spoke other Philippine languages in addition to Tagalog.

TABLE V. The Chong sample words. Target vowels are underlined

Modal (Tone 1)		Breathy (Tone 3)	
kəcaat	"kind of fishing bird"	caat	"ground lizard"
kaap	"road"	rəkaap	"bamboo"
kətaak	"peas"	taak	"water"

TABLE VI. The Mpi sample words

Laryngealized		Modal		Tone	Number of iterations
mi	"to close the eyes"	mi	"to name"	(High falling)	4
ni	"to hurry"	ni	"to awaken"	(Low rising)	4
ti	"muddy"	ti	"to pass over"	(Mid rising)	3
si	(a man's name)	si	"to die"	(High falling)	5
si	"to be dried up"	si	"to be putrid"	(Low rising)	5
si	"to smoke"	si	"to roll a rope"	(Mid rising)	5

3.3. *Chong*

Chong is a Mon-Khmer language with about 500 speakers in Thailand and 5000 in Cambodia (Grimes, 1996). The four "tones" of Chong are distinguished by vowel phonation contrasts. Tone 1 is a level tone produced with modal phonation on a middle pitch. Tone 2 is similar to tone 1, but has a somewhat higher pitch and ends with laryngealization. Tone 3 is a falling tone with breathy phonation. Tone 4 is similar to tone 3, but has a somewhat higher pitch and ends with laryngealization. Only tones 1 and 3 were used in this study.

Table V shows the Chong sample words, with the target vowels underlined. Theraphan Thongkum recorded the four male and four female speakers of the Krathing dialect in December 1986. All speakers were between ages 50 and 60 at the time of the recording. Each word was spoken once in isolation.

3.4. *Mpi*

Mpi is a Tibeto-Burman language spoken by about 2000 people in the villages of Phrae and Phayao in northern Thailand (Grimes, 1996). It is similar to Piyo and Hkatu, which are spoken in China. All Mpi speakers are fluent in Northern Thai. Those who have attended school also know Standard Thai (Bradley, 1991).

There are eight vowel qualities and six tones in Mpi. Contrastive modal and laryngealized phonation occurs on all six tones. (A minimal set of 12 words illustrating these contrasts on the syllable [si] can be heard on *Sounds of the World's Languages* (1991) from the UCLA Phonetics Laboratory.) As in Mazatec, each of the phonation types may also be nasalized.

Table VI shows the Mpi sample words, recorded in April 1976 by James Harris and Peter Ladefoged as part of a longer word list. The single speaker (male) said each word several times in the course of illustrating contrastive tones and phonation types. There

are 3 tokens of [tị] and [ti]; 4 of [mị], [mi], [nị], and [ni]; and 5 of each of the [sị] and [si] words. No words were given in a frame sentence.

The Mpi sample cannot be analyzed statistically with the other samples since it has only one speaker and there are no words with the vowel [a]. The sample set is well balanced internally, however. Although the initial consonants [m, n, s, t] can influence vowel harmonics—breathiness from [s], nasal zeros from [m] and [n]—the problem was minimized by employing pairs of words that have identical initial consonants. For example, [mị] and [mi] can be expected to exhibit similar nasal influences on the vowel; thus, any differences observed between the two vowels should indicate actual phonation parameters.

4. Procedure

The experiment used nine [a]-vowel words from Mazatec and nine from Tagalog. In each language, the nine words consisted of three words from each phonation type. There were 12 speakers of each language, giving 36 tokens from each language and phonation type. Two Mazatec and five Tagalog tokens were eliminated due to faulty recordings.

Corroborating data were obtained from earlier recordings of Chong (three breathy and three modal vowels, eight speakers, for a total of 24 tokens in each phonation type) and Mpi (26 modal and 26 laryngealized vowels, one speaker). From Mpi there were 9 high falling, 8 mid rising, and 9 low rising tokens of each phonation type.

The voiced portion of the target vowels was tagged at 25 ms intervals. The criteria for vowel onset and offset were as follows, in order of priority:

- Vowel starts at an obvious burst and ends at an obvious closure on the spectrogram.
- Vowel must be voiced and show F_1 and F_2 on the spectrogram. Voiced portions without a clear F_1 and F_2 were excluded.
- The Tagalog intervocalic [ʔ], though audible, was not always discernible on the spectrogram. In such cases, an energy reading was made and the point of lowest energy was designated as the onset of the target syllable.
- The first tag was 25 ms after the onset. The last tag could be no closer than 15 ms to the offset.

A fast fourier transform (FFT) was calculated over a 25.6 ms window centered at each tag. Thus, the span of the first window commenced about 12 ms after the onset of vowel formants. The FFT used a Hamming window, 1024 points, and no zero padding. Amplitudes of H_1, H_2, and the highest visually evident harmonic within the F_2 peak (henceforth referred to simply as F_2) were recorded from the FFT. Use of the highest harmonic in F_2 as a surrogate for F_2 can introduce errors, since the harmonic is not usually at the peak frequency of the formant. But the relatively large number of tokens contributing to the mean helps compensate for the problem. $H_1 - H_2$ and $H_1 - F_2$ differences from the FFT were used as measures for comparing phonation type, along with the cepstral peak calculated over the same 25.6 ms window at each tag.

For any given parameter (e.g., $H_1 - H_2$), the data for a single vowel uttered by one speaker were plotted as a line graph with the parameter as the Y-axis and time as the X-axis. Durations of the target vowels in the 12 instances of a single word had standard deviations of 11–34 ms in Tagalog, 21–43 ms in Mazatec. The durations across speakers

were similar enough to allow the graphs for each word to be combined, with each point representing the mean for all speakers at that time window. The standard error of the mean at each point, an estimate of how well the value represents the entire population of speakers, is in the range of 2–3 dB. Points comprising fewer than 12 of the 36 samples (i.e., at the end of the vowel, when vowels of various durations have been averaged together) were excluded, since they do not adequately represent the whole sample.

Since in this set of words the Mazatec modal tokens are shorter than the breathy tokens, it is impractical to compare the phonation types across their time course window by window. Instead, nonmodal vowels at each window were compared to a reference value that was the average of the first two windows of the modal vowel. To facilitate comparison, the same method was used on the other languages as well.

5. Results

The results will be presented in three sections, $H_1 - H_2$, $H_1 - F_2$, and cepstral peak data. Within each section, the Tagalog data will be presented first, followed by Mazatec and Chong. Within each language, modal vowel data will be compared to data from the breathy or after-h vowel set, then to data from the laryngealized or after-ʔ set. Since there was only one Mpi speaker, the Mpi data will not be merged with the results from the other languages, but summarized at the end of the results section.

Throughout the article, statistical differences are based on two-tailed Student's t tests and a significance level of 0.01. Probability values for each time point appear in tables accompanying the figures. All figures use the same numeric scale to facilitate comparison; their ranges are:

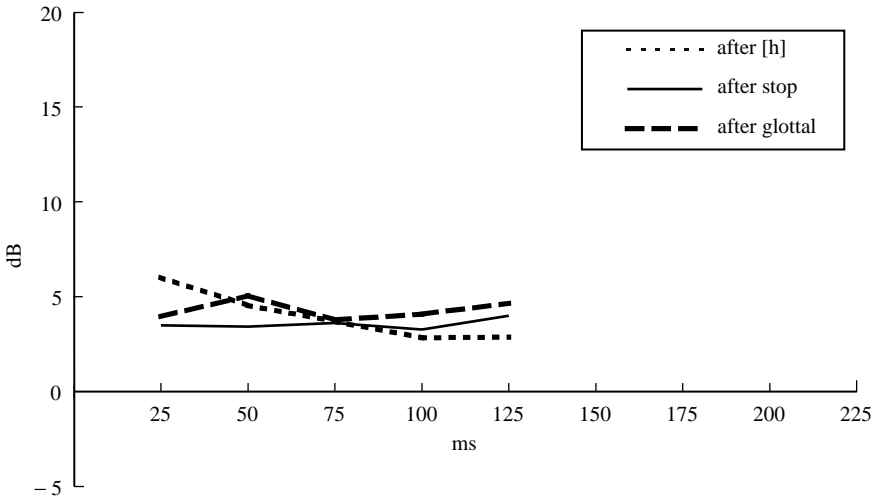|	Parameter	Range
X-axis:	Time	0–225 ms
Y-axis:	H_1–H_2	− 5 to 20 dB
	H_1–F_2	− 5 to 20 dB
	Cepstral peak	2–7 dB

5.1. $H_1 - H_2$

When the glottal cycle has a large open quotient, the amplitude of H_1 is large relative to H_2. Thus, $H_1 - H_2$ should be large during breathy phonation, intermediate for modal phonation, and small for laryngealized phonation, if these distinctions are produced by a variation in open quotient.

5.1.1. *Tagalog*

For convenience, I will refer to vowels following an initial voiced stop as the "stop group", those after [h] as the "h group", and those after [ʔ] as the "glottal group".

Figure 3 shows the 12-speaker mean $H_1 - H_2$ for the three phonation groups in Tagalog. T tests at each time interval show no significant differences between the h group and the reference value (probability values are in the table accompanying the figure). The larger difference between the h group and the stop group at 25 ms suggests an influence of [h] on the vowel prior to 25 ms.

ms	25	50	75	100	125
[h] group *vs.* reference value	0.08	0.4	0.9	0.7	0.7
[ʔ] group *vs.* reference value	0.7	0.3	0.8	0.7	0.4

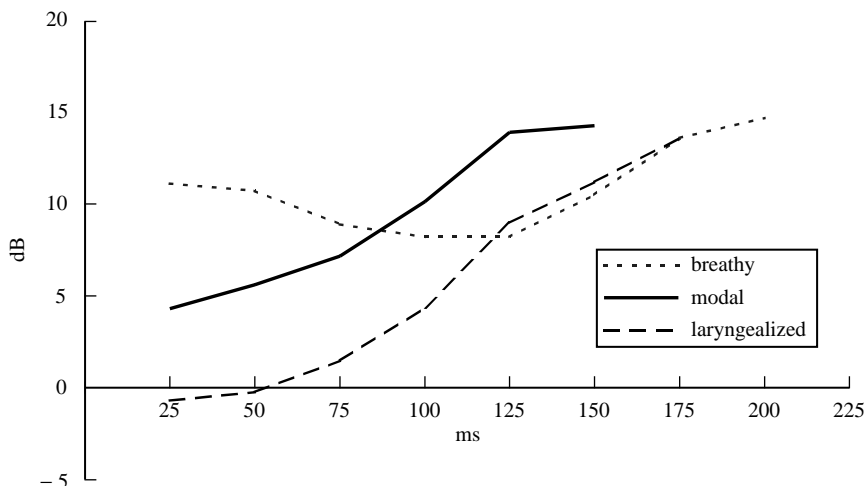Probability values from *t* tests at each time interval

Figure 3. Tagalog. $H_1 - H_2$. Chart shows 12-speaker averages for vowels after [h], stop, and [ʔ]. Table gives probability values of the differences between the post-[h] or postglottal vowels at each time window and the reference value (the average of the poststop vowel values at the first two windows). There are no significant differences.

Likewise, there are no significant differences between the reference value and $H_1 - H_2$ for the glottal group. The Tagalog glottal stops are extremely short, often only the tap of a single glottal pulse, although all were audible stops. Apparently, the stop does not change the position of the vocal folds enough to have a strong effect on $H_1 - H_2$ in the following vowel. It does briefly affect the $H_1 - F_2$ measure, however, as can be seen in Fig. 7.

5.1.2. *Mazatec*

Fig. 4 shows the 12-speaker $H_1 - H_2$ means for Mazatec modal, breathy, and laryngealized vowels. The breathy vowels are significantly different from the modal reference value at all time intervals (probability values are in the table accompanying the figure). The laryngealized vowels are significantly different from the modal reference value except at the 100 ms window. Due to the typical Mazatec breathy offset for utterance-final vowels, the $H_1 - H_2$ values of all three phonation types increase to 15 dB during the last 100 ms. Since the modal vowels in the sample set are of shorter duration, their $H_1 - H_2$ value increases earlier than that of the nonmodal vowels, but the trajectories are nearly identical in shape. Thus apparent differences between nonmodal and modal vowels after 100 ms are the result of disparate vowel durations.

A comparison of the Tagalog and Mazatec $H_1 - H_2$ means shows the nature of the control exercised by Mazatec speakers. Vowels in the Tagalog phonation categories do

Figure 4. Mazatec. $H_1 - H_2$. Chart shows 12-speaker averages for breathy, modal, and laryngealized vowels. Table gives probability values of the differences between the breathy or laryngealized vowels at each time window and the reference value (the average of the modal vowel values at the first two windows).

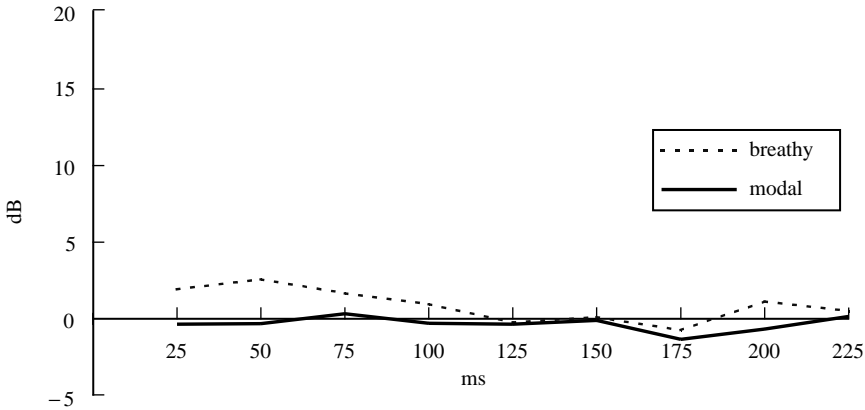| Probability values from t tests at each time interval						
ms	25	50	75	100	125	150
Breathy *vs.* reference value	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
Laryngealized *vs.* ref. value	< 0.01	< 0.01	< 0.01	0.6	< 0.01	< 0.01

not differ significantly. By contrast, Mazatec $H_1 - H_2$ remains significantly different between the phonation categories for at least 100 ms.

The curious time course of the breathy vowels is easily visible in this figure. Where other utterance-final Mazatec vowels show a continuous increase in $H_1 - H_2$ throughout the vowel, the breathy vowels begin with a large $H_1 - H_2$, then adjust toward a more modal value at 75–125 ms, corresponding to the more modal portion of the spectrogram observed by Silverman *et al.* (1995).

5.1.3. *Chong*

Fig. 5 shows the 8-speaker $H_1 - H_2$ means for Chong modal and breathy vowels. While the modal vowels maintain a level value of $H_1 - H_2$ through most of the vowel, the breathy vowels begin with a larger value, and then return to the same range as the modal vowels. T tests at each time interval show no significant differences between the breathy vowels and the modal reference value (probability values are in the table accompanying the figure).

The time course of the $H_1 - H_2$ contrast between breathy and modal is similar in Chong and Mazatec, but the magnitude of the contrast is not the same. Both languages display a maximum contrast for the first 50 ms (not significant in Chong), diminishing to no contrast by the 125 ms measurement, about half way through the vowel. The $H_1 - H_2$ difference between modal and breathy in Chong is never more than 3 dB, however, whereas the maximum difference in Mazatec is over 6 dB. Since the two Chong

Probability values from t tests at each time interval									
ms	25	50	75	100	125	150	175	200	225
Breathy *vs.* ref. value	0.1	0.05	0.2	0.4	0.9	0.7	0.7	0.3	0.6

Figure 5. Chong. $H_1 - H_2$. Chart shows 8-speaker average for breathy and modal vowels. Table gives probability values of the differences between the breathy vowel at each time window and the reference value (the average of the modal vowel values at the first two windows).

categories also differ in tone, there is less need for a large contrast in phonation. In Mazatec, however, each phonation type can occur on any of the three tones, making it necessary to differentiate the phonation types more strongly.

An important difference between Chong and Mazatec is the overall setting. $H_1 - H_2$ of the Chong modal vowels is in the range of Mazatec laryngealized vowel onset. $H_1 - H_2$ of the Chong breathy vowels is in between that of the Mazatec modal and laryngealized vowel onsets. Thus, both Chong phonation types may have a smaller open quotient than their Mazatec counterparts. The more laryngealized setting in Chong is plainly audible in the recordings.
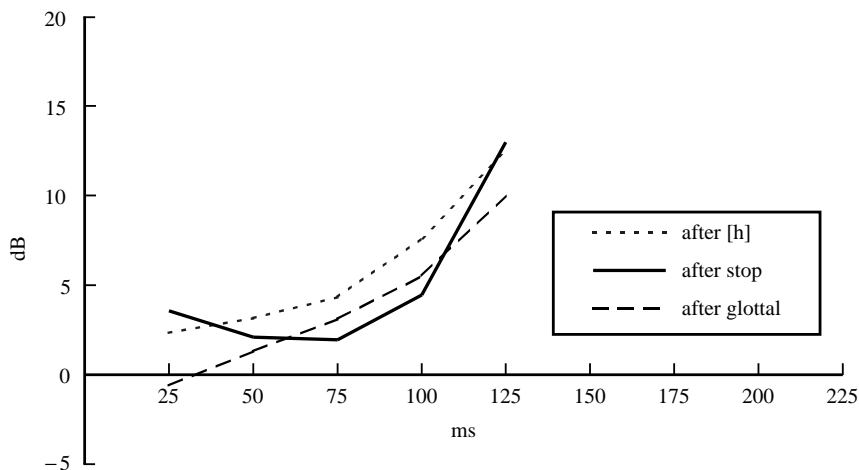
5.2. $H_1 - F_2$

Using the $H_1 - F_2$ amplitude difference as a surrogate for spectral slope, we expect $H_1 - F_2$ to be largest during breathy phonation and smallest—even negative in some instances—during laryngealized phonation.

5.2.1. *Tagalog*

Fig. 6 shows the 12-speaker mean $H_1 - F_2$ of the three Tagalog phonation groups. T tests indicate no significant differences between the nonmodal groups and the modal reference value except at the vowel offset (probability values are in the table accompanying the figure).

Between the glottal group and the reference value, there appears to be some difference at 25 ms. While the glottal stop has no significant effect on $H_1 - H_2$ within the following vowel (Fig. 3), it may cause some leveling of the spectral slope ($H_1 - F_2$) just at the onset

Figure 6. Tagalog. $H_1 - F_2$. Chart shows 12-speaker averages for vowels after [h], stop, and [ʔ]. Table gives probability values of the differences between the nonmodal phonation groups and the reference value. There are no significant differences.

of the following vowel. None of the measured differences in $H_1 - F_2$ is significant, however.
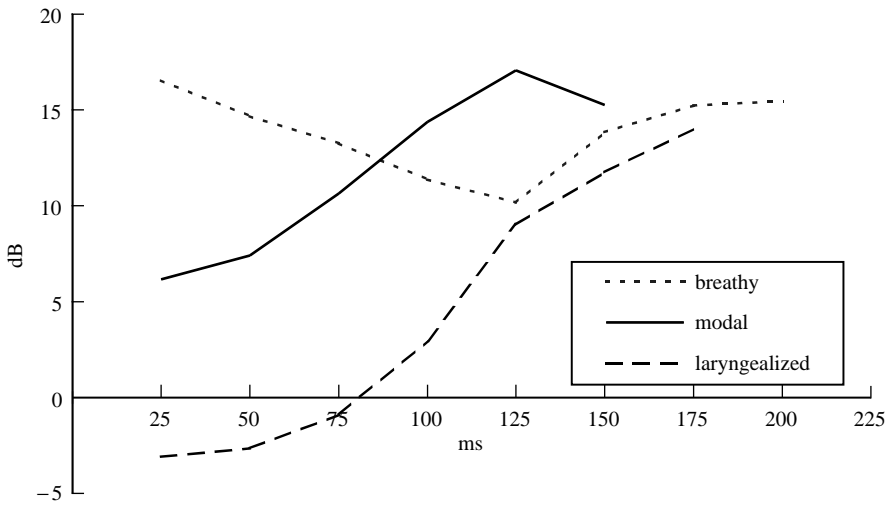
5.2.2. *Mazatec*

Fig. 7 shows the 12-speaker mean $H_1 - F_2$ for the Mazatec modal, breathy, and laryngealized vowels. *T* tests indicate that the breathy vowel differs significantly from the modal reference value for the first 50 ms, and the laryngealized vowel for 75 ms (probability values are in the table accompanying the figure). $H_1 - F_2$ results are similar to those for $H_1 - H_2$. Laryngealization produces smaller dB differences in both $H_1 - H_2$ and $H_1 - F_2$. Breathiness produces larger dB differences on both measures. The differences between phonation categories are noticeably larger in Mazatec than in Tagalog.

Where other utterance-final Mazatec vowels show a gradual steepening in spectral slope throughout the vowel, the breathy vowels begin with a steep slope (a 17 dB difference between F_2 and H_1), level toward a more modal setting (a 10 dB difference) at about 125 ms, and end with a steep spectral slope like the other vowels. As in Fig. 4, apparent differences after 100 ms are a result of the unequal vowel durations.
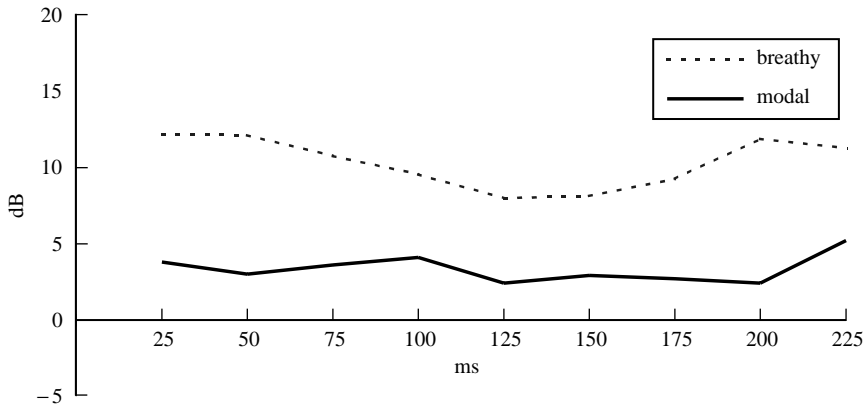
5.2.3. *Chong*

Fig. 8 shows the 8-speaker mean $H_1 - F_2$ for Chong modal and breathy vowels. The two phonation types are significantly different except at the 125 ms window (probability values are in the table accompanying the figure). This is unlike the Mazatec contrast, where breathy and modal vowels differ only for the first half of the vowel.

Probability values from *t* tests at each time interval

ms	25	50	75	100	125	150
Breathy *vs.* reference value	< 0.01	< 0.01	< 0.01	0.025	0.092	< 0.01
laryngealized *vs.* ref. value	< 0.01	< 0.01	< 0.01	0.06	0.29	0.05

Figure 7. Mazatec. $H_1 - F_2$. Chart shows 12-speaker averages for breathy, modal, and laryngealized vowels. Table gives probability values of the differences between the nonmodal vowels and the reference vowel.

Probability values from *t* tests at each time interval

ms	25	50	75	100	125	150	175	200	225
Breathy *vs.* ref. value	< 0.01	< 0.01	< 0.01	< 0.01	0.03	0.01	< 0.01	< 0.01	< 0.01

Figure 8. Chong. $H_1 - F_2$. Chart shows 8-speaker averages for breathy and modal vowels. Table gives probability values of the differences between the breathy vowel at each time window and the reference value.
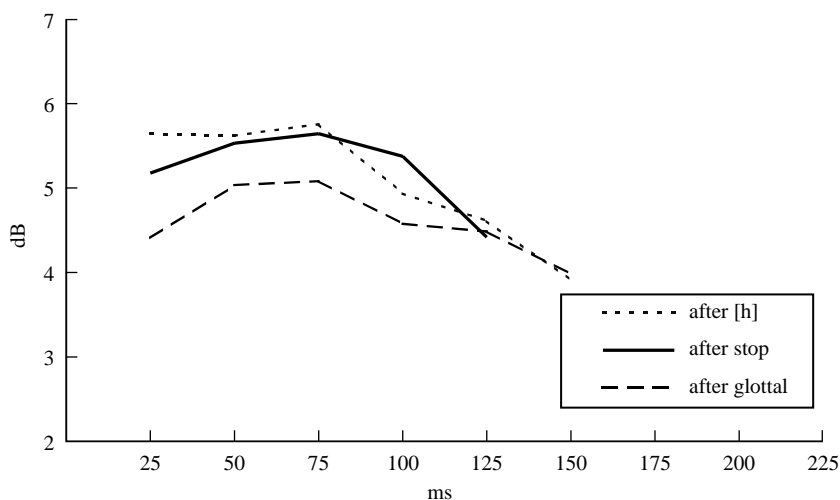
The dB values of both $H_1 - H_2$ and $H_1 - F_2$ are less during the initial portion of the vowel in Chong than in Mazatec. The modal vowels at onset have an $H_1 - F_2$ difference of 4 dB in Chong, 6 dB in Mazatec. The breathy vowels have an $H_1 - F_2$ difference of about 12 dB in Chong, 16 dB in Mazatec.

5.3. *Cepstral peak prominence*

Cepstral peak prominence, the difference in amplitude between the peak cepstral value and the mean of all cepstral values, was used as a measure of periodicity. A larger difference implies a greater ratio of periodic to aperiodic sound in the signal. The pilot study showed this measure to be a good discriminator between the breathy and modal vowels of two Mazatec speakers, but not between laryngealized and modal vowels.
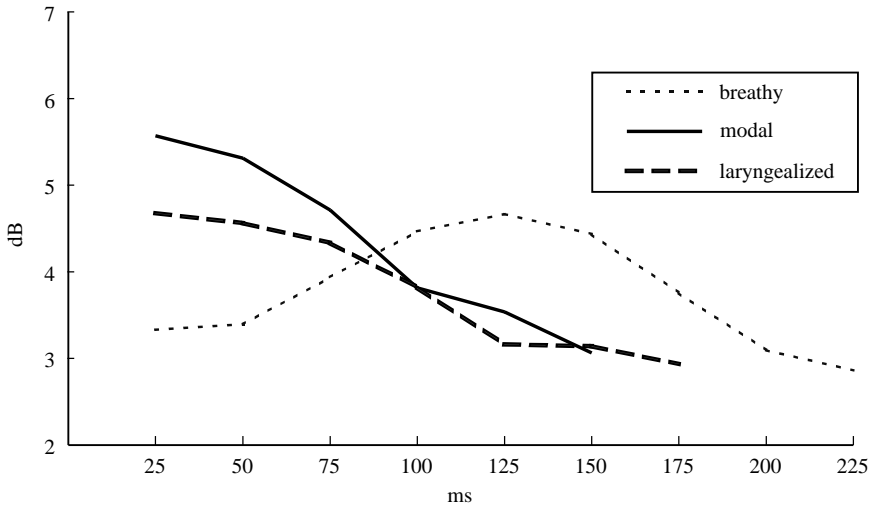
5.3.1. *Tagalog*

Fig. 9 shows the 12-speaker mean cepstral peak prominences of the three Tagalog phonation groups. The cepstral peaks of vowels in the h and glottal groups do not vary significantly from the modal reference value except in the first and last window of the glottal group (probability values are in the table accompanying the figure). Thus, Tagalog vowels display no significant increase in aperiodicity after [h], except a slight increase for as long as 25 ms after [ʔ].

Probability values from *t* tests at each time interval					
ms	25	50	75	100	125
[h] group *vs.* reference values	0.4	0.4	0.3	0.2	0.04
[ʔ] group *vs.* reference values	< 0.01	0.3	0.4	0.03	< 0.01

Figure 9. Tagalog. Cepstral peak prominence. Chart shows 12-speaker averages for vowels after [h], stop, and [ʔ]. Table gives probability values of the differences between the nonmodal phonation groups and the reference value. There are no significant differences.

Probability values from *t* tests at each time interval						
ms	25	50	75	100	125	150
Breathy *vs.* reference value	< 0.01	< 0.01	< 0.01	0.02	0.05	0.02
Laryngealized *vs.* ref. value	0.07	0.04	< 0.01	< 0.01	< 0.01	< 0.01

Figure 10. Mazatec. Cepstral peak prominence. Chart shows 12-speaker averages for breathy, modal, and laryngealized vowels. Table gives probability values of the differences between the nonmodal vowels and the reference value.
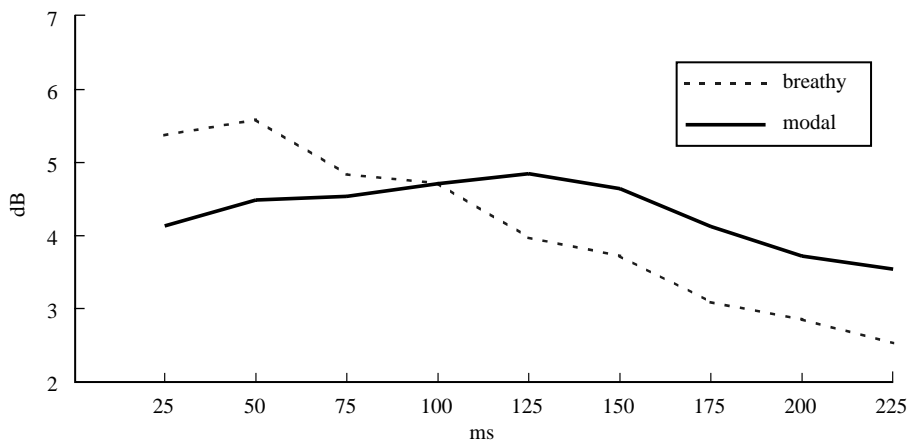
5.3.2. *Mazatec*

Fig. 10 shows the 12-speaker mean cepstral peak prominences of the three Mazatec phonation types. The breathy vowels are significantly different from the modal reference value for the first 75 ms, about half of the vowel (probability values are in the table accompanying the figure). As in earlier measurements, apparent differences after 100 ms result from disparate vowel durations. Laryngealized vowels follow a pattern very similar to that of modal vowels. The significant difference in probability values after 50 ms is an artifact of the analysis method, which uses only the first two windows of the modal vowel. Periodicity is probably not a factor in the contrast between modal and laryngealized vowels.

Breathy vowels are less periodic than modal vowels for about 75 ms. Breathy vowels have the lowest cepstral peaks and laryngealized vowels have intermediate peaks, but both types are less periodic than the modal vowels. After the 50th ms, modal and laryngealized vowels have similar peak values and continue to decline in periodicity. The phonologically breathy vowels, however, become *more* periodic after the 50th ms. Increased periodicity may be an additional means of rendering f_0 more perceptible for distinguishing tone on breathy syllables, as suggested in Silverman (1995).

5.3.3. *Chong*

Fig. 11 shows the 8-speaker mean cepstral peak prominences of Chong breathy and modal vowels. The visible differences on the chart are not significant (probability values are in the table accompanying the figure).

Figure 11. Chong. Cepstral peak prominence. Chart shows 8-speaker averages for breathy and modal vowels. Table gives probability values of the differences between the breathy vowel at each time window and the reference value.

ms	25	50	75	100	125	150	175	200	225
	Probability values from t tests at each time interval								
Breathy *vs.* ref. value	0.02	< 0.01	0.2	0.3	0.4	0.1	< 0.01	< 0.01	< 0.01

It is curious that at onset, breathy vowels appear to be more periodic than modal vowels. A rapid pitch excursion can perturb the cepstral analysis, resulting in lower peaks. But the pitch of the modal set is stable; pitch excursion is not likely to be a factor in its reduced cepstral values.

It may be therefore that in Chong, the *direction* of change is more perceptually salient than the onset value. Thus a breathy vowel is characterized not by being less periodic at onset, but by *becoming* less periodic: there is a steady increase in aspiration noise through the course of the vowel.

5.4. *Mpi*

The Mpi data set had only one speaker and 8–9 tokens of each tone and phonation type, a sample too small and varied for statistical analysis. Nonetheless, certain trends are apparent. Table VII summarizes the durations over which the $H_1 - H_2$, $H_1 - F_2$, and cepstral peak prominence of Mpi laryngealized vowels differed significantly from modal vowels. Wide pitch excursions obstructed comparison of the laryngealized values to a modal reference value as in the other languages. Instead, each laryngealized value was compared to the modal value in the same time window, since the two classes of items had matching tone contours and similar durations.

The primary parameter differentiating modal and laryngealized phonation is the spectral slope ($H_1 - F_2$). The amplitude of F_2 relative to H_1 is 5–6 dB greater for modal vowels than for laryngealized vowels of the same tone. In high-tone vowels, the difference is significant for the first 100 ms. In low- and mid-tone vowels, the difference is significant through nearly the entire vowel.

TABLE VII. Durations over which expected laryngealized values were observed for three tones in Mpi. Durations are given both in ms and as an approximate percentage of the entire vowel

Expected characteristic	High falling tone	Mid rising tone	Low rising tone
$H_1 - H_2$ less in laryngealized vowel than in modal	225 ms (60%)	175 ms (33%)	No effect
$H_1 - F_2$ less in laryngealized vowel than in modal	100 ms (22%)	425 ms (85%)	475 ms (95%)
Cepstral peak prominence less in laryngealized vowel than in modal	No effect	No effect	No effect

TABLE VIII. Durations over which expected breathy values were observed in Tagalog, Mazatec and Chong. Durations are given both in ms and as an approximate percentage of the entire vowel

Expected characteristic	Duration in Tagalog (*allophonic*)	Duration in Mazatec (*phonemic*)	Duration in Chong (*phonemic*)
$H_1 - H_2$ greater in breathy vowel than in modal	No effect	150 ms (100%)	No effect
$H_1 - F_2$ greater in breathy vowel than in modal	No effect	75 ms (50%)	225 ms (100%)
Cepstral peak prominence less in breathy vowel than in modal	No effect	75 ms (50%)	Last 75 ms (final 33%)

The high-tone vowels are also strongly differentiated by $H_1 - H_2$ for about 60% of their duration, whereas these differences last for only 33% of the mid-tone vowels. The low-tone vowels show no $H_1 - H_2$ effect.

Periodicity does not appear to be a factor in making the distinction between laryngealized and modal phonation.

5.5. *Summary of results*

Table VIII summarizes the durations over which an expected characteristic occurred in breathy vowels (Mazatec, Chong) or vowels following [h] (Tagalog.) Durations are given both in ms and as a percentage of the entire vowel. All three parameters, increased $H_1 - H_2$, increased $H_1 - F_2$, and diminished cepstral peak, are exploited in Mazatec to achieve the breathy vowel contrast. Speakers maintain the $H_1 - H_2$ contrast for the entire vowel, and the other contrasts for 50% of the vowel duration.

The Chong breathy vowel contrast depends primarily on increased $H_1 - F_2$ throughout the vowel and diminished cepstral peak amplitude during the final third of the vowel. There is a non-significant increase in $H_1 - H_2$ during the first half of the vowel.

Tagalog speakers produce vowels that have modal values on all parameters by the 25th ms after an [h]. Thus, the human time requirement for resetting glottal parameters appears to be shorter than 25 ms in this environment. The average f_0 for these Tagalog vowels was 140 Hz for males and 200 Hz for females. The return to modal values at 25 ms translates to about 3.5 cycles for males and 5 cycles for females.

TABLE IX. Durations over which expected laryngealized values were observed in Tagalog and Mazatec. Durations are given both in ms and as an approximate percentage of the entire vowel

Characteristic	Duration in Tagalog (*allophonic*)	Duration in Mazatec (*phonemic*)
$H_1 - H_2$ less in laryngealized vowel than in modal	No effect	75 ms* (50%)
$H_1 - F_2$ less in laryngealized vowel than in modal	No effect	75 ms* (50%)
Cepstral peak prominence less in laryngealized vowel than in modal	25 ms (20%)	No effect

*Indicates characteristics that persist through the entire vowel, but are not statistically significant after the 75th ms due to the analysis method. Mpi durations are summarized separately (Table VII) in order to group them by tone.

Table IX summarizes the durations over which an expected characteristic occurred in laryngealized vowels (Mazatec) or vowels following [ʔ] (Tagalog). Here, the phonological contrast in Mazatec is characterized by differences in $H_1 - H_2$ and $H_1 - F_2$, but not in cepstral peak prominence. Mazatec laryngealized vowels are about as periodic as the modal vowels.

Tagalog speakers produced vowels with modal $H_1 - H_2$ and $H_1 - F_2$ values by the 25th ms after a glottal stop. The time required for the articulators to return to their canonical mode of vibration after a glottal stop must be less than 25 ms, although there was still some aperiodic noise, reflected in a reduced cepstral peak, at the 25th ms.

Viewing Tables VII and IX together, the languages displayed no significant cepstral peak difference between modal and laryngealized or post [ʔ] vowels, except for the perturbation at 25 ms in Tagalog. Both $H_1 - H_2$ and $H_1 - F_2$ are used to differentiate modal from laryngealized vowels in both Mazatec and Mpi, except that no $H_1 - H_2$ contrast was apparent on the Mpi low rising tone. One must assume that the low first formant of the [i] vowel interfered with at least some of the H_1 and H_2 amplitude measurements in Mpi. Since the durations of the Mpi contrasts varied with tone contour, one could also speculate that f_0 frequency places constraints on glottal control.

6. Discussion

The introduction posed two questions about the differences between languages where phonation type on vowels is contrastive and languages where it is not. This section will discuss as to how the results have answered those questions.

1. Is nonmodal phonation of *longer duration* in languages with contrastive phonation type?
2. Is nonmodal phonation *more different* from modal phonation in languages with contrastive phonation type? For example, is the abduction for breathy vowels greater in languages where breathiness is a phonological cue?

I will address two additional questions in this discussion.

3. Can phonation type contrasts in all languages be captured by the same acoustic cue?
4. Can different typical values for the measures across languages be unified in terms of a "laryngeal setting continuum" comparable to the VOT continuum?

6.1. *Duration*

On those parameters that are exploited to make a contrast, the duration of the interval where nonmodal values were observed is consistently longer in the contrastive languages, both in absolute time and as a percentage of the complete vowel. (See Tables VII, VIII, and IX.) Between the contrastive languages, however, the duration of nonmodal values and the choice of parameters are not the same. There is apparently no optimum duration for a nonmodal phonation type, so long as it lasts long enough to be perceptible.

The shortest durations of *noncontrastive* [h] and [?] effects on vowels in Tagalog can instruct us about *articulatory* limits. Since the effects last less than 25 ms on most measures, we know that the articulators can return to a modal vibration pattern in less than 25 ms. Longer durations measured in other languages would be particular to the speaker or language, and not due to human articulatory limitations. The swift return to modal values in Tagalog could also indicate that modal is not simply a default position of the glottis for vowels, but is phonologically specified in Tagalog.

The shortest durations of *contrastive* values on vowels in Mazatec, Chong, and Mpi can instruct us about *perceptual* limits and about the interplay between articulatory effort and perceptual salience. In these data, few of the nonmodal values persist through the entire vowel. There are several possible explanations for the result.

- It may require extra effort to sustain a nonmodal configuration of the vocal folds.
- Surrounding segments may make conflicting articulatory demands, a possibility in the $H_1 - H_2$ data for Chong breathy vowels, which are followed by voiceless stops.
- There may be conflicting perceptual demands, as Silverman (1995) postulated for Mazatec breathy vowels, where contrasts in both tone and phonation type must be perceptible on the same vowel.

With additional measures and a larger number of subjects, this study confirms Silverman's observation that Mazatec breathy vowels are breathy for a limited time. In Figs 4, 7, and 10, the breathy vowel averages show a continuous trend toward modal values through 125 ms, after which they return toward breathy values at the pre-pausal offset. There is no articulatory requirement that the breathy parameters be of short duration: the Chong data, for example, show us that a spectral slope value typical of breathy voice can be maintained through the entire vowel. Therefore, a perceptual explanation like Silverman's for Mazatec is more plausible. Chong does not need to include a modal portion in breathy vowels, because hearing the pitch is not crucial. Although Chong is a tone language, each tone associates with a distinct phonation type. The perception of phonation types in the absence of pitch information would be adequate to discern the tone distinctions.

In Mazatec, only the breathy vowels adjust toward modal; laryngealized vowels follow a steady course from start to finish (compare the slopes of the phonation types in Figs 4, 7, and 10). This fact also supports Silverman's theory that the adjustment in breathy

vowels is driven by perceptual requirements. Since laryngealized phonation has a strong periodic component, there is no need to adjust the phonation in order to enhance pitch information.

Interestingly, Cho, Jun & Ladefoged (2002) have found that nonmodal phonation persists for about half the duration of vowels following Korean aspirated and fortis stops. Although as in Tagalog the phonation difference is not contrastive on vowels, the robustness of the differences suggests that voice quality on the following vowel is important for perceiving the three-way contrast of Korean stops.

6.2. *Magnitude of difference*

Tables X and XI summarize the magnitude of the absolute differences between modal and nonmodal phonation for the measures used in this study. Table X shows the differences between the modal vowel all-speaker average and the breathy vowel all-speaker average at each time frame through 125 ms. Table XI shows the differences between modal and laryngealized vowels through 125 ms.

At time frames where a parameter supports a phonological contrast, the modal–nonmodal differences are larger in the contrastive language than in Tagalog. The differences are not larger at time frames where a parameter is not performing contrastively: the Mazatec breathy vowels as they transition to modal values at 100 and 125 ms (Table X, all measures), Chong breathy vowels at 25 and 50 ms where phonation is still periodic (Table X, cepstral peak difference), and cepstral peak difference on laryngealized vowels (Table XI).

6.3. *Choice of acoustic cues*

With the variety of acoustic cues to phonation type, it would not be surprising to discover that different languages would exploit different cues. Some experimental evidence shows that this is the case.

TABLE X. Magnitude of breathy minus modal difference on vowel measures at 25 ms intervals*

ms	25	50	75	100	125
$H_1 - H_2$ difference (dB)	(Expected relationship: breathy > modal)				
Mazatec	7	5	2	− 2	− 6
Chong	2	3	1	1	0
Tagalog	2	1	0	− 1	− 2
$H_1 - F_2$ difference (dB)	(Expected relationship: breathy > modal)				
Mazatec	10	7	3	− 3	− 7
Chong	8	9	7	5	6
Tagalog	− 1	1	2	3	1
Cepstral peak difference	(Expected relationship: breathy < modal)				
Mazatec	− 2	− 2	− 1	1	1
Chong	1	1	0	0	− 1
Tagalog	0.5	0	0	− 0.5	0

*Negative numbers indicate that breathy value was less than modal value.

TABLE XI. Magnitude of modal minus laryngealized difference on vowel measures at 25 ms intervals*

ms	25	50	75	100	125
$H_1 - H_2$ difference (dB)	(Expected relationship: modal > laryngealized)				
Mazatec	5	6	6	6	5
Tagalog	0	-2	0	-1	-1
$H_1 - F_2$ difference (dB)	(Expected relationship: modal > laryngealized)				
Mazatec	9	10	12	11	8
Tagalog	4	1	-1	-1	3
Cepstral peak difference	(Expected relationship undetermined)				
Mazatec	1	1	0	0	0
Tagalog	1	0	1	1	0

*Negative numbers indicate that modal value was less than laryngealized value. Mpi data are not included for comparison, since those samples used a different vowel. The Mpi results do in fact agree with those for Mazatec.

In perceptual tests, Klatt & Klatt (1990) found that amplitude of aspiration noise was the dominant factor in judgments of breathiness by 11 American English speakers listening to synthesized vowels. Hillenbrand, Cleveland & Erickson (1994) found that cepstral peak, another possible indicator of aspiration noise, was the most important predictor of breathiness ratings from 20 American speakers listening to natural speech stimuli.

On the other hand, Bickley (1982) found no correlation between the amount of aspiration noise and judgments of breathiness by six native Gujarati speakers listening to synthesized Gujarati vowels modeled on natural speech. Instead, breathiness ratings correlated with increased $H_1 - F_2$. The results of Ladefoged & Antoñanzas-Barroso (1985) agree with those of Bickley and contradict those of Hillenbrand, and Klatt and Klatt. Breathiness judgments by 10 American speakers listening to breathy and modal vowels by 10 !Xõo speakers correlated more strongly with the spectral slope measure $H_1 - F_1$ than with aspiration noise. In agreement with these two studies, Gobl & Ní Chasaide (1999) report competent breathy/modal discrimination on a synthesized [a] vowel when either the high spectrum (the synthesizer parameter TL) or the low spectrum (a combination of the synthesizer parameters OQ, SQ, and the bandwidths of the first two formants) is varied. Changes in aspiration noise (synthesizer parameter AH) did not serve to discriminate between breathy and modal vowels in the absence of TL changes.

Different languages appear to prefer different cues for breathiness. The feature "breathy" encompasses a suite of source parameters, which include reduced vocal fold closure duration (possibly resulting in a larger $H_1 - H_2$ difference), more gradual vocal fold closure (possibly resulting in a steeper spectral slope), and a glottal chink or vocal fold abduction to allow air leakage and frication through the glottis. One member of the suite can be the favored parameter in a language, although all of them are present to some degree. Thus in one language, the percept of breathiness may be produced with a loosely adducted glottis and higher open quotient, resulting in a larger $H_1 - H_2$, whereas in another language, aspiration noise generated at a posterior opening may be superimposed on more or less modal vibration at the adducted anterior portion, with $H_1 - H_2$ being the same as for modal voice. Similarly, for laryngealization a language could employ changes in spectral slope, $H_1 - H_2$, or both.

Although this study has no perceptual data to verify which parameters are salient to listeners in the selected languages, the acoustic data show which ones are produced most in each language. On average, all humans are equally adept at controlling parameters: differences between languages are not due to articulatory limitations. Thus if a language tends to favor one parameter, the preference may be based on acoustic or perceptual requirements. The preferred parameter for the breathiness *vs.* modal contrast in Chong is spectral slope (more sloped for breathy than for modal), whereas Mazatec uses all the parameters about equally. The preferred parameter for making the laryngealized *vs.* modal contrast in Mpi is also the spectral slope (more level for laryngealized than for modal), while Mazatec uses both spectral slope and H_1–H_2. It should be recalled here that laryngealized phonation in Mpi also includes other strongly audible factors (possible faucal tension, velarization, or tongue lowering) that have not been addressed in this study.

Given that languages achieve phonation contrasts by different means, it is important for phonation studies to employ more than one measure in order to discern the contrasts and characterize them faithfully.

6.4. *The laryngeal continuum*

Lisker and Abramson (1964) investigated cross language differences in the timing of voice onset relative to consonant release, revealing that such seemingly unrelated consonant distinctions as voiced/voiceless, aspirated/unaspirated, and some cases described as fortis/lenis are all categories on the single articulatory continuum of voice onset time (VOT), within which each language has from one to three contrastive categories. The locations of the category boundaries on the VOT continuum are not the same from language to language. For example, Puerto Rican Spanish /b/ has an average VOT of −138 ms in isolated words (p. 392), while American English /b/ has an average of 1 ms (p. 394), which is in the same range as the Spanish /p/. (VOT is measured relative to consonant release. A negative VOT indicates that voicing began prior to release.) Thus, an English listener whose perceptual system is set to English categories may perceive Spanish /p/ as /b/. Language-specific perceptual categories cause the continuum of VOT to be chunked into blocks. Listeners perceive categorically; they have difficulty in distinguishing between two items that are within the same category in their native language.

The current study illuminates another such continuum, which is the degree of laryngeal tension. (Laryngeal tension is a term of convenience. This paper does not attempt to determine whether the continuum is based on longitudinal tension, abduction and adduction, or some other factor.) The concept of laryngeal tension as a continuum was introduced in Ladefoged (1971), where it is referred to as the glottal stricture continuum. More recently the topic received detailed treatment in Ladefoged & Maddieson (1996). A language like English, which makes no phonation-type contrast on vowels, has only one phonological category on the glottal stricture continuum. An analogy on the VOT continuum would be a language like Maori, with only one series of plosive stop consonants (Maddieson, 1984, p. 345). Many languages have two categories of vowels on the glottal stricture continuum, which include Chong and Mpi in this study. A two-way stricture distinction is often the basis of register differences in Southeast Asian languages that have two vowel registers. Some languages such as Mazatec have three categories of vowels on the continuum. Better recognition of the glottal stricture continuum could

TABLE XII. Male $H_1 - H_2$ averages (in dB) for each phonation type

Phonation type	Mazatec	Chong
Breathy	9	-1
Modal	4	-3
Laryngealized	1	

TABLE XIII. Male $H_1 - F_2$ averages (in dB) for each phonation type

Phonation type	Mazatec	Chong
Breathy	16	6
Modal	9	0
Laryngealized	5	

lead to a more unified view of vowel registers, breathy and creaky vowels, Korean consonant contrasts, and other nonmodal phenomena on both consonants and vowels.

As with VOT categories, locations of category boundaries on the laryngeal continuum vary from language to language. The phonation categories of the sample languages illustrate this point. Table XII compares the male $H_1 - H_2$ averages during the first half of the vowel in the combined tokens from Mazatec and Chong. (These vowels are as nearly comparable as possible. Except for one of the three Mazatec words, all vowels begin on mid-tone.) Differences in $H_1 - H_2$ may simply be due to differences in recording conditions, or they may indicate actual category locations on the phonation continuum. All three Mazatec categories have positive values for $H_1 - H_2$. Chong values are negative. A Chong breathy vowel might be heard as a laryngealized vowel by a Mazatec speaker, comparable to the confusion where VOT categories overlap between Spanish and English.

Table XIII compares the male $H_1 - F_2$ averages during the first half of the vowel. These form a similar pattern to the $H_1 - H_2$ averages. On this measure as well, a Chong breathy vowel might be heard as a laryngealized vowel by a Mazatec speaker.

7. Concluding summary

This research has examined the duration of nonmodal phonation and the magnitude of its difference from modal phonation. Measures of $H_1 - H_2$, $H_1 - F_2$, and cepstral peak prominence showed that the difference between nonmodal and modal vowels is of greater duration and magnitude in languages where nonmodality is contrastive on vowels.

The results showed that the vocal folds can return from [ʔ] or [h] to modal phonation in less than 25 ms, but that it is possible to sustain nonmodal phonation through an entire vowel. The study provides further evidence that source parameters such as open quotient and spectral slope can be controlled independently of each other. The several

acoustic cues to phonation type may be produced separately or in combination; languages do not all use the same set of cues.

Finally, this research has provided evidence for a continuum of glottal stricture. Among the languages with a vowel phonation contrast, comparable categories did not have the same spectral characteristics from language to language, but within each language the categories were consistent across speakers. This fact indicates that from the continuum of possible spectra, each language establishes its own range of spectra for phonation categories.

References

Bickley, C. (1982) Acoustic analysis and perception of breathy vowels, *Speech Communication Group Working Papers*, Research Laboratory of Electronics, MIT, **1,** 73–83.

Blankenship, B. (1997) *The time course of breathiness and laryngealization in vowels.* Doctoral dissertation, University of California at Los Angeles.

Bradley, D. (1991) *Thai hill tribes phrasebook.* Victoria, Australia: Lonely Planet Publications.

Childers, D. G. & Lee, C. K. (1991) Vocal quality factors: analysis, syntheses, and perception, *Journal of the Acoustical Society of America*, **90,** 2394–2410.

Cho, T., Jun, S. & Ladefoged, P. (2002) Acoustic and aerodynamic correlates to Korean stops and fricatives, *Journal of Phonetics*, **37,** 193–227.

Cohn, A. C. (1990) *Phonetic and phonological rules of nasalization.* Doctoral dissertation, University of California at Los Angeles.

Gobl, C. & Ní Chasaide, A. (1988) The effect of adjacent voiced/voiceless consonants on the vowel voice source: a cross-language study, *Speech Transmission Laboratory Quarterly Progress and Status Report*, **2–3,** 23–59.

Gobl, C. & Ní Chasaide, A. (1999) Perceptual correlates of source parameters in breathy voice. *Proceedings of the XIVth International Congress of Phonetic Sciences*, pp. 2437–2440.

Grimes, B. F. (editor) (1996) In *Ethnologue.* (R. S. Pittman & J. E. Grimes, consulting editors). Dallas, TX: Summer Institute of Linguistics, Inc. http://www.ethnologue.com/web.asp

Hanson, H. (1997) Glottal characteristics of female speakers: acoustic correlates, *Journal of the Acoustical Society of America*, **101,** 466–481.

Hillenbrand, J., Cleveland, R. A. & Erickson, R. L. (1994) Acoustic correlates of breathy vocal quality, *Journal of Speech and Hearing Research*, **37,** 769–778.

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

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

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

Ladefoged, P. (1983) The linguistic use of different phonation types. In *Vocal fold physiology. Contemporary research and clinical issues* (D. M. Bless & J. H. Abbs, editors), pp. 351–360. San Diego: College-Hill Press.

Ladefoged, P. (2001) *Vowels and consonants.* Oxford: Blackwell Publishers. (Accompanying audio and visual illustrations are also available at http://hctv.humnet.ucla.edu/departments/linguistics/VowelsandConsonants/vowels/chapter12/ chapter12.html).

Ladefoged, P. & Antoñanzas-Barroso, N. (1985) Computer measures of breathy phonation, *UCLA Working Papers in Phonetics*, **61,** 79–86.

Ladefoged, P. & Maddieson, I. (1996) *The sounds of the world's languages.* Oxford: Blackwell Publishers Inc.

Laver, J. (1980) *The phonetic description of voice quality.* Cambridge: University of Cambridge Press.

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

Lisker, L. & Abramson, A. S. (1964) A cross-language study of voicing in initial stops: acoustical measurements, *Word*, **20,** 384–422.

Löfqvist, A. & McGowan, R. S. (1992) Influence of consonantal envelope on voice source aerodynamics, *Journal of Phonetics*, **20,** 93–110.

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

Ní Chasaide, A. & Gobl, C. (1997) Voice source variation. In *The Handbook of phonetic science* (W. J. Hardcastle & J. Laver, editors), pp. 427–461. Oxford: Blackwell Publishers Inc.

Pierrehumbert, J. & Talkin, D. (1992) Lenition of /h/ and glottal stop. In *Papers in laboratory phonology II. Gesture, segment, prosody* (G. J. Docherty & D. R. Ladd, editors), pp. 90–119. Cambridge: Cambridge University Press.

Schachter, P. (1987) Tagalog. In *The world's major languages* (B. Comrie, editor), pp. 933–958. London and Sydney: Croom Helm.

Silverman, D. (1995) *Phasing and recoverability*. Doctoral dissertation, University of California at Los Angeles (Also in outstanding dissertations in linguistics series. New York: Garland Publishing, 1997).

Silverman, D., Blankenship, B., Kirk, P. & Ladefoged, P. (1995) Phonetic structures in Jalapa Mazatec, *Anthropological Linguistics*, **37**, 70–88.

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

Stevens, K. N. (1999) *Acoustic phonetics*. Cambridge, MA: MIT Press.

Stevens, K. N. & Hanson, H. M. (1995) Classification of glottal vibration from acoustic measurements. In *Vocal fold physiology. Voice quality control* (O. Fujimura & M. Hirano, editors), pp. 146–170. San Diego, CA: Singular Publishing Group Inc.

Titze, I. R. (1995) Definitions and nomenclature related to voice quality. In *Vocal fold physiology. Voice quality control* (O. Fujimura and M. Hirano, editors), pp. 335–342. San Diego, CA: Singular Publishing Group Inc.

UCLA Phonetics Laboratory (1991) *Sounds of the world's languages*. Hypercard stacks. Los Angeles: University of California. (Ordering information at http://www.linguistics.ucla.edu/faciliti/sales/software.htm).