The time course of
breathiness and laryngealization in vowels

Barbara Blankenship

Abstract

In a language where breathiness or laryngealization is a contrastive property of vowels, such non-modal phonation lasts longer and may be of greater magnitude than in a language where it is an accident of consonant context. 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 the difference between nonmodal and modal vowels is stronger and lasts longer in Mazatec than in Tagalog. Contrary to expectations, cross-speaker variation is not greater in Tagalog.

The main experiment examined words from 6 male and 6 female speakers of each language. A second experiment used 4 male and 4 female speakers of Chong, 1 male speaker of Mpi, and 1 male and 10 female speakers of Navajo. Three breathy, 3 laryngealized, and 3 modal vowels from each speaker were analyzed. To determine the time course of phonation effects, measurements were made at 25 ms intervals through each vowel. The measurements were H1-H2 (hypothesized to reflect the open quotient of the glottal vibration), H1-F2 (an approximation of spectral slope, hypothesized to reflect the abruptness of vocal fold closure), and cepstral peak prominence (a measure of periodicity).

Table of contents

List of figures

List of tables

Acknowledgment

This dissertation is at least half Don’s. Not only did he take up all the slack in running the household, raising the kids, and maintaining the income – the usual duties of a graduate student’s spouse – but he also consulted on the experimental design, methodology, and statistics.

To attempt graduate studies after you have a job, children, and a mortgage requires three elements: a patient, supportive family, an extremely tolerant faculty, and a motive.

I entered graduate school as the children were entering first grade. Now 11th graders, they have managed with only half a mother all these years. Thanks to them for doing without, and to Don for being both mother and father.

The UCLA linguistics department has dealt uncomplainingly with the complexities of fitting a part time student into a full time program. I was privileged to study with such bright and generous teachers as Bruce Hayes, Pat Keating, Peter Ladefoged, Ian Maddieson, Bruce Derwing, Pam Munro, and Bob Stockwell. The research was nourished by lively discussions at the phonetics lab, a place where everyone is welcome and everyone brings ideas. Peter and Jenny, thank you for establishing its open, positive environment. Pat, thank you for keeping it that way, and for providing us every year with new tools.

Thanks also to my tireless committee — Patricia Keating, Sun-Ah Jun, Jody Kreiman, and Peter Ladefoged — for months of coaching and support, especially to Pat, who provided the idea, gave me enough rope, and then told me when it was time to stop.

And what of the motive? Mine is founded on an unspoken dare, planted unawares by my father, who both scorned and envied academic pursuit. Mom is still here to cheer me on, but it was Dad, long since passed away, who continued to needle me. Thanks to you both.

1. Introduction

Studies of the linguistic uses of voice source variation have dealt in some detail with the time course of fundamental frequency (F0) both as it relates to tone in tone languages and to phrasal pitch patterns in non-tone languages. F0 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. See figure 107 at the back of this book for photographs of a glottis producing the three types.

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. (An example of a non-creaky laryngealized vowel is shown in figure 44.) In this study, all laryngealized vowels will be considered as a single phonation type, whether they are creaky or not. Titze (1995) prefers the term “pressed voice” for this type of phonation. Pierrehumbert and Talkin (1992, 93) refer to a “braced configuration” of the vocal folds. See also Laver (1980) for a discussion of distinctions between laryngealized and creaky.

The time course of variations in phonation type has recently received some attention. Silverman *et al.* (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 for the remainder. It is surprising that non-modal 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.)

Languages that do not have phonation specification on vowels can have breathy vowels near [h] or aspirated consonants, and laryngealized vowels near [/] or laryngealized consonants. What do we know about the duration of non-modal phonation in vowels where phonation type is not contrastive? Lofqvist and

MacGowan (1992) found that vowels after [h] or aspirated $[p^h]$ have breathy phonation during approximately the initial 7 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 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 Lofqvist and MacGowan's non-contrastive vowels. (Since the Swedish and English vowel durations were not given, we cannot determine what 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 the research designs of the two studies. The designs are summarized in table 1.

Table 1.
Comparison of research designs

	LOFQVIST & MACGOWAN (1992)	SILVERMAN ET AL (1995)
Speakers	1 female, English 1 male, Swedish	6 female, Mazatec 6 male, Mazatec
Samples	1 stressed nonsense syllable in frame, 12 samples per speaker	8 isolated Mazatec monosyllabic words, 1 sample per speaker
Vowels	a	i ae a o u
Average vowel duration	not reported	255.6 ms
Basis for analysis	open quotient of airflow wave	narrow band spectrogram

Is contrastive breathiness about twice as long as non-contrastive breathiness across languages? Does the duration of contrastive breathiness vary little across speakers, and that of non-contrastive breathiness vary more, as it appears to in the difference between the two speakers in Lofqvist and MacGowan's study? Is the duration of breathiness in Mazatec similar to that of other languages that have contrastive breathiness?

A central thesis of Silverman (1995) is that simultaneous phonological features that would tend to obscure each other perceptually may be realized non-simultaneously in order to maximize the salience of each one. Thus the Mazatec breathy vowel is breathy during the first half but becomes modal during the second half in order to render F0 more perceptible for distinguishing tone. The harmonics are weak in breathy phonation, making it more difficult to determine pitch. Since the harmonics are strong in laryngealized phonation, tone perception of laryngealized vowels should be less of a problem. Thus one can predict that Mazatec laryngeal 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 affects the time course of that phonation during a vowel. Both breathy and laryngealized phonation will be considered, in languages where they are contrastive and in languages where they are not contrastive but arise as the result of a preceding [h] or [/]. The study will focus on three questions.

I. Is non-modal phonation of longer duration in languages with contrastive phonation type?

II. Is non-modal phonation more consistent across speakers in languages with contrastive phonation type?

III. Is non-modal phonation more different from modal phonation in languages with contrastive phonation type?

The primary goal of the study is to add to the body of knowledge about sounds of the world's languages, in an area that has not been much studied within linguistic phonetics. Cross-language differences in the timing of vowel onset relative to consonant release were investigated fruitfully by Lisker and Abramson (1964), revealing that such seemingly unrelated consonant distinctions as voiced/voiceless, aspirated/unaspirated, and some cases described as fortis/lenis are all part of the single articulatory continuum of voice onset time (VOT). It is possible that glottal tension is another phonetic continuum upon which languages establish categories for the distinctions between breathy, modal, and laryngealized segments. Further data on the timing of glottal tension will give a better understanding of the linguistic uses of this continuum, leading to a unified theory of the articulatory parameters that correlate with phonation distinctions on vowels and consonants, as well as their use in ejectives and implosives.

A secondary goal is to provide further measures for describing the dimensions of phonation in normal voices. Knowledge of these dimensions has applications in clinical assessment, speech synthesis, and automated speech recognition. If several languages that do not have a phonation contrast on vowels display a similar time course of nonmodal phonation after [h] or [/], that may indicate an articulatory constraint on how rapidly the vocal folds can return to their canonical mode of vibration. For the languages that do have a phonation contrast, a knowledge of contrast durations can inform us about perceptual limits and about the kinds of articulatory control available to achieve perceptual contrasts.

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 breathiness or laryngealization. A speaker may use one or more articulatory adjustments for each phonation type. Table 2 summarizes some of the articulations posited for breathy vowels in Hanson (1996), Stevens & Hanson (1995), Klatt & Klatt (1990), Gobl & Ni Chasaide (1988). The right column lists articulatory maneuvers, the middle column gives the postulated effect of each articulation on the glottal vibration, and the right column lists the acoustic outcome claimed to result from the maneuver.

There is still little experimental evidence for the presumed correlation between the articulatory events listed in the left column of the table and the acoustic outcomes listed in the right column. One goal of the study reported in Holmberg *et al.* (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 (H1-H2) 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.

Hanson (1997), in a study of American English gender differences in breathiness, found that H1-H2 does not correlate with measures of spectral slope (H1-F1 and H1-F3.) Thus spectral slope is related to an aspect of articulation other than open quotient. A widely held theory that originated with Stevens (1977) maintains that spectral slope correlates with the abruptness or gradualness of vocal fold closure. While there are no articulatory observations yet to support the theory, it does provide a convenient framework for discussion. Throughout this study I shall assume that H1-H2 correlates with open quotient and spectral slope correlates with abruptness of vocal fold closure.

Table 3 summarizes possible articulations for laryngealized vowels. The items are purely speculative, but often based on the theories summarized in Table 2.

Figure 1, from Klatt & Klatt (1990) illustrates some of the acoustic effects of posited breathy and laryngealized settings in synthesized voices. The left column represents posited laryngealized phonation. The circle picture at the top is a schematic of the glottis, pressed closed (the small triangles at the bottom represent the arytenoids). The waveform below is the pattern of glottal opening. The glottis is

Table 2.
Elements of breathy articulation

ARTICULATION	EFFECT ON GLOTTAL VIBRATION	ACOUSTIC RESULT
Vocal folds remain closed for a shorter time	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 F0, thus H1 has markedly higher amplitude than the other harmonics (large H1-H2 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 F1 bandwidth (but bandwidth can also increase due to absorption by the walls of the vocal tract).
Arytenoids held open, allowing airflow between the arytenoids even when the main portions of the vocal folds are closed	Wave includes both periodic and aperiodic elements	Non-periodic aspiration noise at high frequencies (conflicts with spectral slope measurement); increased F1 bandwidth

Table 3.
Possible elements of laryngealized articulation

ARTICULATION	EFFECT ON GLOTTAL VIBRATION	ACOUSTIC RESULT
Folds may remain closed for a longer time	Smaller ratio of open phase to complete cycle (open 25-45% of cycle — Childers & Lee 1991)	H1 amplitude often less than that of H2.
Glottis closes suddenly, exciting frequencies throughout the spectrum	Glottal vibration approaches closed phase abruptly	Spectral slope above 2000 Hz is more gradual, not as steep.
Glottis opens more slowly than it closes	Glottal cycles have a skewed shape	Spectral slope above 2000 Hz is more gradual, not as steep.
Glottal closure at irregular time intervals	Glottal vibration has unequal cycles	Jitter

open for only about 1/3 of the cycle; at the end of the open phase, the folds close abruptly. The spectrum at the bottom of the column shows two characteristics that can be used to distinguish laryngealized phonation: the first harmonic (H1) has lower amplitude than the second (H2), and above about 1500 Hz, the harmonics remain stronger than those of modal phonation shown in the middle column.

The middle column represents modal phonation. The glottis is open for about half of the cycle and does not close as abruptly. Therefore the spectrum has higher amplitude at the lower harmonics, and drops off at a rate of about 12 dB per octave. In this study I shall refer to the drop-off rate as “spectral slope,” rather than the alternate terms “roll off” or “tilt”.

The right column shows breathy phonation. The folds close gradually from front to back, but due to the position of the arytenoids, the glottis never closes completely. The source wave therefore is almost sinusoidal, producing a spectrum dominated by the low harmonics that drops off at a rate of about 18 dB per octave. The spectrum also shows aspiration noise above 1500 Hz, due to air flowing through the opening between the arytenoids while the glottis is otherwise closed.

In actual breathy speech, the speaker may produce a spectrum with a steep slope, aspiration noise, or both. Since we do not know which strategy a speaker will use, it is desirable to have measures of each of the possible acoustic results in order to make sure of capturing the contrast.

2.2. Considerations in selecting measurements

A great variety of acoustic measures have been used by scientists to determine vocal source characteristics. The choice of measure is determined in part by the goal of the study and by the kind of data available.

Clinical speech scientists describe source characteristics that correspond to perceived vocal pathology. (For recent reviews of this research area, see Rabinov *et al* 1995, and Bangayan *et al* 1997.) The data often are isolated vowels of several seconds' duration, longer than in natural speech. They are uttered by speakers whose acoustic voice characteristics are known. Common measures include jitter (perturbations in duration of the glottal cycle), shimmer (perturbations in amplitude of the glottal cycle), and the relationship of harmonic to non-harmonic elements.

Scientists concerned with speech synthesis study vocal source variations in order to develop more natural sounding synthesizers. For purposes of modeling, the phonation of an individual speaker is quantified by a number of measures of the shape of the first differential of the glottal waveform. Results of the calculations are used to set synthesizer parameters to produce a similar waveform. The synthetic waveform is then subjected to perceptual and other tests to assess its success in imitating a human speaking voice. As in clinical speech research, the samples are often of longer duration than in natural speech.

While these two research groups are concerned with source characteristics that distinguish individual speakers, linguists are concerned with source characteristics that are similar for all speakers of a given language or dialect, namely those that are exploited as phonological distinctions or for sociolinguistic purposes. The samples, usually from fluent speech, are very brief, and measures that require calculations over a long duration are inappropriate. Typically only audio recordings are available, limiting analysis to acoustic measures.

2.3. Methods selected for trial

The data used in this study impose two constraints on the choice of analysis method. (a) Only tape recordings were available; therefore only acoustic measures could be used. (b) 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 made it impractical to select measures that would require a great deal of manipulation. A pilot study was performed to test methods.

Ten measures were selected for testing. The first uses the relative amplitudes of the fundamental and its next harmonic, H1-H2. This measure is thought to correlate with open quotient. The next eight measures, comparing the amplitude of the fundamental with that of another formant, are ways of approximating the spectral slope. The last measure, cepstral peak amplitude, indicates how much the signal is periodic. No measures of jitter, harmonics to noise ratio, or bandwidth were made, since an earlier pilot study had shown these to be unreliable in windows of this duration.

2.3.1. Relationship between amplitudes of the fundamental and its first harmonic: H1-H2; (Average of H1and H2) - F1

The difference between the amplitudes of the fundamental frequency (H1) and its first harmonic (H2), has been used primarily to distinguish between breathy and modal phonation (Bickley 1982, Huffman 1987, Klatt & Klatt 1990). When the vibration of the vocal folds has a large open quotient, the spectrum of a vowel is dominated by the fundamental frequency. The fundamental (H1) has a markedly higher amplitude than the other harmonics, resulting in a large positive value for H1-H2. Breathy vowels typically show these large values, while modal and laryngealized vowels have small positive or even negative values. Frequently the H1-H2 value is greater for modal than for laryngealized phonation in comparable vowels from the same speaker.

Hanson (1997) determined that H1-H2 does not correlate strongly with F0-F1, F0-F3, or harmonics-to-noise measures. Thus it is likely that the parameter denoted by H1-H2 is independent of the other glottal parameters.

2.3.2. Relationship between amplitudes of the fundamental and another formant: H1-F1; H1-F2; H1-F3; F2-F3; F2 - Max (F3, F4, F5); Range 1 - Range 2; Range 2 - Range 3

Some studies have found correlations between phonation type and the prominence of H1 relative to the vowel formants F1 or F2 (Ladefoged 1983; Kirk, Ladefoged & Ladefoged 1984; Gobl & Ni Chasaide 1992). Instead of H1, Stevens (1988) compared the mean of H1 and H2 to F1.

Stevens & Hanson (1995) and Hanson (1996) used the amplitude difference of H1-F3 as an indicator of spectral slope. Klatt & Klatt (1990) compared the amplitude difference of F2-F3. An alternate measure compared F2 to the amplitude of F3, F4, or F5, whichever was greatest . If these measures correlate successfully with spectral slope, they will give small values for laryngealized vowels, medium values for modal, and large values for breathy. In breathy vowels this measure can be invalidated by high frequency aspiration noise.

Hammarberg *et al.* (1986) divided the long-term average spectrum into three frequency ranges: 0-2000 Hz, 2000-5000 Hz, and 5000-8000 Hz, as illustrated in figure 2. The differences between the peak amplitudes in each of these ranges give an indication of spectral slope. For breathy vowels they found a large difference between the first and second ranges (henceforth, R1 and R2) relative to modal vowels, but little difference between the second and third ranges (R2 and R3) because of high frequency noise. One would expect these measures to give smaller differences for laryngealized vowels, where the spectrum is less sloped.

2.3.3. Periodicity: cepstral peak

Several recent studies have used cepstral peak prominence, a measure of aperiodicity, as an indicator of aspiration noise in the vowel. Hillenbrand *et al.* (1994) evaluated thirteen acoustic measures on how well they correlated with American English listeners' ratings of perceived breathiness. The study included H2-H1, several measures of spectral slope, and measures of aperiodicity. Periodicity measures accounted for about 80% of the variance in breathiness ratings. The most successful of these was cepstral peak prominence. Klatt and Klatt (1990) had also found that listener judgments of breathiness were more strongly influenced by aspiration noise than by H1 amplitude, a finding opposite to that of Fischer-Jorgensen (1967) and Bickley (1982).

A cepstrum is an inverse spectrum generated by taking the FFT (Fast Fourier Transform) of the log magnitude values of a power spectrum. The spectrum of a highly periodic signal shows well defined harmonics; its cepstrum has a prominent peak at a location corresponding to the duration of the F0 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.

Figure 3 shows the cepstrum of two signals of differing periodicity. The upper left illustration is the spectrum of a modal vowel with a fundamental frequency of about 150 Hz. In the lower left is the corresponding cepstrum with a prominent peak at about 6.6 ms, which is the cycle duration of F0. On the right are the spectrum and cepstrum of a breathy vowel, with an F0 of about 125 Hz and a weak cepstral peak at 8 ms. To normalize for differences in energy, Hillenbrand *et al.* take the difference between the cepstral peak and a linear regression of the cepstral values. For ease of calculation I have used the mean of the cepstral values in place of a regression. Assuming that 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

2.4. Methods trial

To determine which analyses distinguished best between phonation types, I tested the ten measures discussed above on a sample data set from Mazatec, a language that contrasts breathy, modal, and laryngealized phonation on vowels. The sample comprised the three words shown in table 4, spoken in isolation by twelve speakers; two of the speakers were recorded twice, giving a total of 14 samples of each word. The words were selected because they have the same vowel quality, and thus the formant locations should be similar within speaker. It was not possible to match the vowels for tone and duration. (All data samples are transcribed in 1993 IPA notation. Raised 1, 2, and 3, indicate low, mid, and high tone; raised letters indicate aspiration or nasalization on a consonant. A colon (:) marks long vowels. There is a dieresis (..) below phonemically breathy vowels and a tilde (~) below phonemically laryngealized vowels.)

Table 4.
Mazatec sample words for the methods trial

^{n}da^{2}	‘good’	(modal)
^{n}da:^{23}	‘hard’	(breathy)
tSa^{3}	‘burden’	(laryngealized)

To avoid effects from the preceding consonant, the first 50 ms of the vowel were omitted from consideration. Each measurement was made on a 25 ms portion starting at the 50th ms. Spectrograms confirmed that the phonation type was canonical at this location. For example, phonemically breathy vowels that start breathy and later become modal are still breathy during the portion sampled. Figure 4 illustrates the location of the test window on a breathy vowel. All measurements were made using a Kay CSL Computerized Speech Lab model 4300B. The results of each measure were compared within speaker to determine which measures differed in the appropriate direction between the nonmodal and the modal samples.

Table 5 shows the number of times each measure succeeded on 14 samples of each phonation type. The expected relationship between the values for laryngealized, modal, and breathy phonation is shown at the top of each group of measures.

Table 5.
Results of the methods trial: number of times each measure succeeded on 14 samples.

MEASURE	LARYNGEALIZED	BREATHY
	Expected:	Expected:
OPEN QUOTIENT	laryngealized < modal	breathy > modal
H1-H2	12	13
(H1+H2)/2 - F1	10	13
	Expected:	Expected:
SPECTRAL SLOPE	laryngealized < modal	breathy > modal
H1-F1	9	13
H1-F2	13	12
H1-F3	13	7
F2-F3	6	3
F2 - (greater of F3 or F4)	7	6
R1-R2	12	3
R2-R3	6	7
	Expected:	Expected:
PERIODICITY	laryngealized < modal	breathy < modal
Cepstral peak	10	14

H1-F2 and H1-F3 were the best measures for distinguishing between laryngealized and modal samples. In agreement with the findings of Hillenbrand *et al.*, cepstral peak prominence best differentiated breathy from modal samples. Two of the measures, H1-H2 and H1-F2, worked well for both kinds of comparison.

The laryngealized samples misidentified by H1-H2 (thought to reflect open quotient), and H1-F2 and H1-F3 (thought to reflect abruptness of vocal fold closure) were not from the same speakers. Likewise the breathy samples misidentified by H1-H2 and H1-F2 were not from the same speakers. Therefore a combination of these measures could be used to discover instances of non-modal phonation in cases where a single measure might not. I selected two measures for the main experiment: H1-H2 as the measure of open quotient and H1-F2 as the measure of abruptness. I used cepstral peak amplitude, the only available measure of periodicity, as well, with the reservation that cepstral peak is not always a useful measure for laryngealized vowels, since aperiodic noise is not a necessary concomitant of laryngealization.

3. Main experiment

3.1. Language materials

The study required one language with contrastive breathiness or laryngeality on vowels, and another language that has only modal vowels but offers consonant environments that induce breathiness or laryngeality 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 [/] 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. There was no control for syllabic position within the word. The Mazatec sample vowels are primarily from monosyllables; the Tagalog sample vowels are usually in the penultimate syllable of a longer word.

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 engender unusual acoustic properties on the vowel (e.g. nasals). There were six male and six female adult speakers of each language.

3.1.1. Jalapa Mazatec

Mazatecan 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 15,500 speakers, 4,600 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 occur on all five vowels and all three tones. Silverman *et al.* (1995) give a more detailed description of the phonetics of Jalapa Mazatec.

The Mazatec sample words are shown in table 6. Tone 1 is low, tone 2 mid, and tone 3 high. The speakers recorded these items as part of a much longer word list illustrating all the phonetic contrasts of Mazatec. Due to the length of the list, words were recorded without a framing sentence. The recordings were made in Jalapa de Diaz in April 1993 by Paul Kirk and Peter Ladefoged. Most of the male speakers were bilingual in Spanish and Mazatec; most of the females were monolingual.

Table 6.
The Mazatec sample words. (Tone 1 is low, 2 is mid, and 3 is high.)

LARYNGEALIZED		MODAL		BREATHY	
Ba ^{2}	‘carries’	^{n}da ^{2}	‘good’	^{n}da: ^{23}	‘hard’
tSa ^{3}	‘load, burden’	na ^{1}	‘woman’	^{n}dja: ^{23}	‘cornflour drink’
^{n}ka^{2}	‘high, tall’	ka ^{2}	‘bald’	kiNa: ^{23}	‘he fastened’

3.1.2. Tagalog

Tagalog is a member of the western Malayo-Polynesian branch of the Austronesian language family. Originally spoken on the southern part of Luzon, it has spread throughout the Philippines since 1937, when it was selected as the national language (Schachter 1987). It is spoken as a first language by 14,850,000 people (Grimes 1996). Tagalog has five modal vowels; lexical stress is marked primarily by increased vowel duration. Among the consonants are both [h] and [/], although no symbol for the latter is included in the Tagalog alphabet: all orthographically adjacent vowels are pronounced with an intervening [/]. Schachter (1987) contains a description of the Tagalog phonetic inventory.

The Tagalog sample words are shown in table 7. Underlines indicate the target vowels. 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. Although words were read once in isolation and once in a frame, I used only the isolated words in this study, to facilitate comparison with the Mazatec set. All twelve readers were fluent in both English and Tagalog. Two were fluent in Spanish and two spoke other Philippine languages in addition to Tagalog.

Table 7.
The Tagalog sample words. (Target vowels are underlined.)

AFTER [/]		AFTER [b/d/g]		AFTER [h]	
pa/a	‘foot’	baga	‘embers’	pahajag	‘article’
taga/ajos	‘manager’	dagat	‘sea’	mahal	‘expensive’
pa/alam	‘good-bye’	gaga	‘fool’(fem.)	mahalai	‘depraved’

3.2. Procedure

The main experiment used nine [a]-vowel words from Mazatec and nine from Tagalog. There were 12 speakers of each language, giving about 100 samples from each. Two Mazatec and five Tagalog samples were eliminated due to faulty recordings or unexpected pronunciations, e.g. Tagalog [pa/a] pronounced as [paa]. 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:

I. Vowel starts at an obvious burst and ends at an obvious closure on the spectrogram.

II. Vowel must be voiced and show F1 and F2 on the spectrogram. Voiced portions without a clear F1 and F2 were not included.

III. The Tagalog intervocalic [/], though audible, was not always discernible on the spectrogram. In such cases, an energy reading was made; I took the point of lowest energy as the onset of the second syllable. An example is shown in figure 5.

IV. The first tag was 25 ms after the onset. The last tag could be no closer than 15 ms to the offset. Figure 6 shows the placement of the tags.

A Fast Fourier Transform (FFT) was calculated over a 25.6 ms window centered at each tag. Amplitudes of H1, H2, and the highest harmonic within the F2 peak (henceforth known simply as F2) were determined from the FFT. H1-H2 and H1-F2 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 measure, (e.g. H1-H2), the data for a single vowel uttered by one speaker were represented as a line graph with the measure 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 point. Figure 7 shows a sample graph for one word with the twelve speakers averaged. Such a graph gives a coherent picture of the progress of phonation through the vowel. 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 three speakers (i.e. at the end of a vowel) were deleted, since they do not adequately represent the whole sample.

3.3. Results

The results will be presented in three sections. The first section discusses H1-H2, assumed to relate to open quotient, the percentage of the glottal cycle during which the glottis is open. The second section covers spectral slope data

(H1-F2), assumed to relate to the abruptness of glottal closure. The third section shows cepstral peak data, which indicate the amount of aperiodic sound in the vowel. Within each section, the Tagalog data will be presented first, followed by the Mazatec data. 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-/ set.

Data on individual words are provided in Appendix A. I will consider only the combined averages (3 words X 12 speakers = 36 samples) in this chapter. Due to individual differences in vowel duration, time frames near the ends of vowels may have fewer than 12 speakers remaining. Time frames with fewer than 12 of the 36 samples are omitted from consideration. Throughout the chapter, statistical differences are based on two-tailed Student's *t* tests and a significance level of 0.01. *P* values for each time point are given in tables accompanying the text, with shaded cells where the value is significant. All figures use the same scale to facilitate comparison; their ranges are:

	Parameter	Range
X axis:	Time	0 to 225 ms
Y axis:	H1-H2	-5 to 20 dB
	H1-F2	-5 to 25 dB
	Cepstral peak	2 to 7 dB

3.3.1. H1-H2 (hypothesized indicator of open quotient)

When the vocal folds remain open for a longer portion of the glottal vibration cycle, the amplitude of H1 is greater relative to H2. Thus H1-H2 should be large during breathy phonation, intermediate for modal phonation, and small for laryngealized phonation, if these distinctions are produced by a variation in glottal closure duration.

3.3.1.1. 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 [/] as the glottal group.

Figure 8 shows the 12-speaker means for stop group and h group. Error bars indicate one standard error of the mean, an estimate of how well these values represent the entire population of speakers. (Due to differences in vowel duration across speakers, the error bars grow very long after 125 ms, where few speakers are represented.) *T*-tests at each time interval show no significant differences, (*p* values are given in table 8), although there is a greater difference at 25 ms than elsewhere. There may be some influence of [h] on H1-H2 of the following vowel prior to 25 ms.

Table 8.
P *values of Tagalog H1-H2 differences*

ms	25	50	75	100	125
[h] group vs. stop group	.08	.4	.9	.7	.4
[/] group vs. stop group	.8	.3	.9	.6	.7

Figure 9 shows the stop vowels and glottal vowels together. Although the figure suggests the possibility of a longer open phase in the stop group, *t*-tests indicate that the means are not significantly different, (*p* values are given in table 8). Thus [/] has no significant effect on duration of glottal closure within the following vowel, when compared to vowels in the stop group. Spectrograms show the Tagalog glottal stops to be of extremely short duration, often only the tap of a single glottal pulse, although they were audible stops. (An example is shown in figure 10.) Apparently such a gesture does not change the position of the vocal folds enough to have much effect on the vocal-fold closure duration of the following vowel. It does briefly affect the H1-F2 measure, however, as I shall discuss with figure 15.

3.3.1.2. Mazatec

Figure 11 shows the 12-speaker means for Mazatec modal and breathy vowels together. Error bars indicate the standard deviation of the mean. Two tailed *t* tests at each time interval show the groups to be significantly different for the first 50 ms, about 33% of the vowel (*p* values are given in table 9, with significant values shaded.) Apparent differences after 100 ms are an effect of the disparate vowel durations. Due to the typical Mazatec breathy offset for utterance-final vowels, the H1-H2 values of all three phonation types ascend to 15 dB during the last 100 ms. Since the modal vowels in the sample set are of shorter duration, they ascend earlier than the breathy vowels, but the final trajectories are nearly identical in shape.

The time course of the breathy vowels is easily visible in this figure. Where other utterance-final Mazatec vowels show a continuous rise in H1-H2 throughout the vowel, the breathy vowels begin with a high H1-H2 for 50 ms, then switch to a more modal setting, ending with a rise like the other vowels.

Figure 12 shows the means for laryngealized and modal vowels together. *T* tests indicate that the groups are significantly different through the first 125 ms, about 83% of the vowel, (*p* values are given in table 9). For most of the vowel, the Mazatec laryngealized vowels have a longer closed phase than do modal vowels.

Table 9.
P *values of Mazatec H1-H2 differences*

ms	25	50	75	100	125	150
breathy vs. modal	.0000003	.0001	.1	.1	.00008	.1
laryngealized vs. modal	.000008	.0000004	.00001	.00002	.005	.01

A comparison of the Tagalog and Mazatec H1-H2 means shows the nature of the control exercised by Mazatec speakers. Vowels in the Tagalog h group have an increased H1-H2 for about 25 ms while the vocal folds return to their normal position after being apart for the [h]. By contrast, Mazatec breathy vowels continue to have an increased H1-H2 for 50-75 ms as the speaker holds the vocal folds at an intermediate distance apart: not as wide as for [h] since that would prevent voicing, but not as narrow as in modal phonation. Vowels in the Tagalog glottal group show no apparent influence on H1-H2 from the brief vocal fold closure of the [/]. By contrast, Mazatec speakers hold the folds closer together through the entire laryngealized vowel. Smaller error bars indicate a greater consistency among Mazatec speakers, another indication of the control exercised to achieve phonemic contrast.

3.3.2. H1-F2 (hypothesized indicator of abruptness of vocal fold closure)

When the vocal folds come together gradually over their length, they excite primarily the lower frequencies of the vocal tract. The resulting sound wave is nearly sinusoidal, dominated by the fundamental frequency. The spectrum of such a wave has a steep downward slope, with most of the energy near F0 and very little energy at higher frequencies. When the folds come together all at once, they excite a wider range of frequencies. The resulting sound wave is more complex, consisting of many frequencies. The spectrum has a more gradual slope, with energy spread across all the frequencies. Figure 13 shows the FFT spectrum at the 50th ms for a Mazatec breathy, modal, and laryngealized [a] vowel spoken by the same speaker. Arrows indicate the second formant, which has a high amplitude in the laryngealized vowel and a low amplitude in the breathy vowel.

Using H1-F2 as a measure of spectral slope, we expect H1-F2 to be largest during breathy phonation, where the vocal folds do not close simultaneously over their entire length. H1-F2 should be smallest — possibly even negative — during laryngealized phonation, where the vocal folds are tense and come together abruptly.

3.3.2.1. Tagalog

Figure 14 compares the spectral slopes of the Tagalog stop group and h group. T tests indicate no significant differences between the groups, (p values are given in table 10). The vocal folds undulate in the same manner after [h] as after a voiced stop.

Between the glottal group and the stop group (figure 15) there appears to be some difference at 25 ms. Thus while the glottal stop has no significant effect on vocal fold closure duration (H1-H2) within the following vowel (figure 9), it may cause the vocal folds to stiffen and meet more abruptly (H1-F2) just at the onset of the following vowel. But none of the measured differences in H1-F2 are significant, (*p* values are given in table 10).

Table 10.
P *values of Tagalog H1-F2 differences*

ms	25	50	75	100	125
[h] group vs. stop group	.5	.5	.2	.06	.9
[/] group vs. stop group	.03	.7	.5	.6	.2

3.3.2.2. Mazatec

Figure 16 shows the 12-speaker means for the Mazatec modal and breathy vowels. *T* tests at each time interval indicate that the groups are significantly different for the first 50 ms, about 33% of the vowel, (*p* values are given in table 11). As in figure 11, apparent differences after 100 ms are an effect of the unequal vowel durations. Where other utterance-final Mazatec vowels show a gradual steepening in spectral slope throughout the vowel, the breathy vowels begin with a steep slope for 50 ms, then level to a more modal setting, ending by steepening again like the other vowels.

Table 11.
P *values of Mazatec H1-F2 differences*

ms	25	50	75	100	125	150
breathy vs. modal	.0000002	.0002	.2	.2	.004	.6
laryngealized vs. modal	.0000005	.0000007	.0000003	.000002	.002	.2

Figure 17 shows the means for laryngealized and modal vowels together. *T* tests indicate that the groups are significantly different for the first 125 ms, about 83% of the vowel, (*p* values are given in table 11). The vocal fold cycle during

Mazatec modal vowels produces a steeper spectral slope than that of laryngealized vowels.

Results of the spectral slope measurements for Mazatec parallel those for H1-H2. Compare figure 16 to figure 11, and figure 17 to figure 12. Laryngealization is produced by a combination of holding the glottis closed longer (as indicated by low values of H1-H2) and stiffening the vocal folds (as indicated by low values of H1-F2). Breathiness is produced by a combination of holding the glottis open longer (high values of H1-H2) and relaxing the vocal folds to meet in a more undulating motion, (high values of H1-F2). Laryngealization is maintained through about 83% vowel. Breathiness is maintained for about 33% of the vowel.

3.3.3. Cepstral peak prominence (periodicity)

The difference in amplitude between the peak cepstral value and the mean of all cepstral values was used as a measure of periodicity. As noted before, 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 breathy and modal vowels in Mazatec, but less reliable between laryngealized and modal vowels.

3.3.3.1. Tagalog

Graphs comparing the cepstral peak amplitude of the Tagalog stop group to those of the h group and glottal group are shown in figure 18 and figure 19. The cepstral peaks of vowels in the h and glottal groups do not vary significantly from those of the stop group, (*p* values are given in table 12). Tagalog vowels do not display an increase in aperiodicity after [h], although a non-significant increase after [/] is visible at 25 ms.

Table 12.
P *values of Tagalog cepstral differences*

ms	25	50	75	100	125
[h] group vs. stop group	.2	.8	.8	.3	.6
[/] group vs. stop group	.04	.2	.1	.06	.9

3.3.3.2. Mazatec

Figure 20 compares the cepstral peak prominences of Mazatec breathy and modal vowels; figure 21 compares laryngealized to modal vowels. The breathy vowels are significantly different from the modal for the first 50 ms, about 33% of the vowel, (*p* values are given in table 13). As in earlier measurements, apparent differences after 100 ms are an effect of the disparate vowel durations. Laryngealized vowels are not significantly different from modals, (*p* values are given in table 13).

Table 13.
P *values of Mazatec cepstral differences*

ms	25	50	75	100	125	150
breathy vs. modal	.0000005	.000006	.08	.13	.01	.009
laryngealized vs. modal	.04	.09	.3	1.0	.4	.9

Breathy vowels are less periodic than modal vowels for 50 ms. Breathy vowels have the lowest cepstral peaks and laryngealized vowels have intermediate peaks, but both types have more aperiodicity 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. Increased periodicity may be an additional means of rendering F0 more perceptible for distinguishing tone on breathy syllables.

3.4. Summary

Table 14 summarizes the durations over which an expected characteristic occurred in breathy vowels (Mazatec) or vowels following [h] (Tagalog). Durations are given both in ms and as a percentage of the entire vowel. All three parameters, open quotient, abruptness of closure, and aperiodic sound, are exploited in Mazatec to achieve the contrast between breathy and modal vowels. The duration of the contrast is controlled by the speaker to last for about half of the vowel duration, after which both types of vowel are modal — according to Silverman (1995), in order to make pitch information more perceptually salient.

Tagalog speakers produce vowels that have modal values on all parameters by the 25th ms after an [h]. Thus the human time requirement for making the requisite gestures appears to be smaller than 25 ms in this environment.

The average F0 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, less than the 8- and 11-cycle durations reported by Lofqvist and MacGowan (1992). Thus Lofqvist and MacGowan’s results may reflect cultural differences rather than a human articulatory limit.

Table 14.
Durations of expected breathy values

Characteristic	Duration in Tagalog *allophonic*	Duration in Mazatec *phonemic*
H1-H2 greater than for modal vowel (increased open quotient)	no effect	50 ms (33%)
H1-F2 greater than for modal vowel (less abrupt vocal fold closure)	no effect	50 ms (33%)
Cepstral peak amplitude less than for modal vowel (increased non-harmonic)	no effect	50 ms (33%)

Table 15 summarizes the durations over which an expected characteristic occurred in laryngealized vowels (Mazatec) or vowels following [/] (Tagalog). Once again, the phonological contrast in Mazatec is reflected by differences in all three parameters. Data suggest that the laryngeal gestures associated with H1-H2 and H1-F2 are controlled by the speaker to last through the entire vowel. Although the table indicates that the cepstral peak amplitude of laryngealized vowels was lower than that of modal vowels for only 25 ms, the similarity between the two phonation types after 25 ms is due to a decrease in modal vowel peaks. The laryngealized vowel peaks remained relatively unchanged throughout the vowel.

Tagalog speakers produced vowels with modal values by the 25th ms after a glottal stop. Thus the time required for the articulators to return to their canonical mode of vibration after a glottal stop must be less than 25 ms. Mazatec speakers make the distinction between laryngealized and modal vowels by a gesture that changes both H1-H2 and H1-F2, but does not alter the amount of harmonic sound in the signal.

Table 15.
Durations of expected laryngealized values

CHARACTERISTIC	DURATION IN TAGALOG *allophonic*	DURATION IN MAZATEC *phonemic*
H1-H2 less than for modal vowel (decreased open quotient)	no effect	125 ms 83%
H1-F2 less than for modal vowel (more abrupt vocal fold closure)	no effect	125 ms 83%
Cepstral peak amplitude less than for modal vowel (increased non-harmonic sound)	no effect	no effect

4. Second experiment

The Tagalog and Mazatec results are useful, but are more interesting if they can be corroborated with data from other languages. The second experiment explores other languages that have some characteristics in common with either Tagalog or Mazatec. The experimental questions and procedures are the same in this section as in the main experiment.

4.1. Language materials

The languages selected for the second experiment 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 laryngealization on vowels, the laryngealization occurs near the end. Since this study concentrates on vowel onsets, I did not include Chong laryngealization.

Mpi was selected as an example of contrastive 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 beyond the scope of this study.

Navajo was selected because it is similar to Tagalog in having both [h] and [/] in its consonant inventory, but no contrastive phonation on vowels.

4.1.1. 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. I used only tones 1 and 3 in this study.

The Chong sample words are shown in table 16, with the target vowels underlined. Four male and four female speakers of the Krathing dialect were recorded in December 1986, by Theraphan Thongkum. All speakers were between ages 50 and 60 at the time of the recording. Each word was spoken once in isolation.

Table 16.
Chong sample words. (Target vowels are underlined.)

MODAL (TONE 1)		BREATHY (TONE 3)	
kEcaat	'kind of fishing bird'	caat	'ground lizard'
kaap	'road'	rEkaap	'bamboo'
kEtaak	'peas'	taak	'water'

4.1.2. Mpi

Mpi is a Tibeto-Burman language spoken by about 2000 people in the villages of Phrae and Phayao in northern Thailand (Grimes 1966). 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 the CD-ROM Sounds of the World's Languages (1991) from the UCLA Phonetics Laboratory.) As in Mazatec, each of the phonation types may also be nasalized.

The Mpi sample words are shown in table 17. These words were 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 samples of ti and ti; 4 of mi, mi, ni, and ni; and 5 of each of the si and si words. No words were given in a frame sentence.

Table 17.
Mpi sample words

LARYNGEALIZED		MODAL		
mi	'to close the eyes'	mi	'to name'	(high falling)
ni	'to hurry'	ni	'to awaken'	(low rising)
ti	'muddy'	ti	'to pass over'	(mid rising)
si	'a man's name'	si	'to die'	(high falling)
si	'to be dried up'	si	'to be putrid'	(low rising)
si	'to smoke'	si	'to roll a rope'	(mid rising)

Since the /mi/, /ni/, and /ti/ pairs had high falling, low rising, and mid rising tones, respectively, I selected /si/ pairs with the same tones for inclusion; the three other /si/ pairs were not used. A sample set thus balanced for both tone and phonation type can show unambiguously the interaction between those two factors, a possibility that was lacking in the Mazatec sample.

The words beginning with /s/ are not ideal for a vowel phonation study, since the open position of the vocal folds during the initial voiceless fricative may influence the phonation of the following vowel. Due to the small number of samples that would result from their omission, they were included in this experiment. I will note cases where the /s/ words gave different results from the other words.

4.1.3. Navajo

Navajo is an Athapaskan language spoken in the Four Corners area of the United States. According to the 1990 U.S. census, there are 148,530 speakers including 7,616 monolinguals, (Grimes 1996).

Navajo has four modal vowels, all of which can be contrastively nasalized. Each vowel occurs in two lengths and two tones. Although the language has /// and /h/, as well as other glottalized and aspirated stops in its consonant inventory, there is no contrastive use of phonation type on vowels.

The Navajo words are shown in table 18. The vowels under study are underlined. The sample included only one example of a voiced stop followed by [a]. Eleven speakers (ten female, one male) were recorded at Monument Valley High school in Kayenta, Arizona, on July 26, 1993 by Joyce McDonough and Martha Austin-Garrison. All of the speakers are bilingual in English and Navajo, but use Navajo daily.

Table 18.
Navajo sample words. (Target vowels are underlined.)

AFTER [/]		AFTER STOP		AFTER [H]	
Gaa/ask/idii	‘camel’	baGaa/	‘wool’	hanahodinilnih	‘talk endlessly’
bika/aniljeed	‘help him’			hanahodiniSniih	‘talk endlessly’
jah/anijood	‘drive them in’			nahaS/na	‘I move, stir’

4.2. Results

The results for each language will be presented in three sections, H1-H2, H1-F2, and cepstral peak. Within each section, modal vowel data will be compared to data from the breathy or after-h vowel set, then to data from the glottal or after-/ set.

Data on individual Chong and Navajo words are provided in Appendix B. Only the combined averages are considered in this chapter. Due to individual differences in vowel duration, time frames near the ends of vowels may have fewer speakers. Time frames with fewer than one third of the sample vowels remaining are omitted from consideration. Throughout the chapter, statistical differences are based on two-tailed *t* tests and a significance level of 0.05. To accommodate the variety of data, figures are scaled differently for each language. Thus the figures cannot be compared visually between languages. Their ranges are:

	PARAMETER	CHONG	MPI	NAVAJO
X axis:	Time	0 to 300 ms	0 to 575 ms	0 to 150 ms
Y axis:	H1-H2	-5 to 10 dB	-20 to 0 dB	0 to 20 dB
	H1-F2	-5 to 15 dB	-15 to 15 dB	0 to 35 dB
	Cepstral peak	2 to 7 dB	2 to 10 dB	2 to 7 dB

4.2.1. Chong

4.2.1.1. H1-H2 (hypothesized indicator of open quotient)

Figure 22 shows the 8-speaker means for modal and breathy vowels. Error bars indicate the standard deviation of the mean. While the modal vowels maintain a level value of H1-H2 through most of the vowel, the breathy vowels begin with a higher value, then return to the same range as the modal vowels. *T* tests at each time interval show no significant differences between the groups (*p* values are given in table 19).

Table 19.

P *values of Chong H1-H2 differences*

ms	25	50	75	100	125	150	175	200	225
breathy vs. modal	.1	.06	.4	.4	.9	.9	.7	.3	.9

The time course of the H1-H2 contrast between breathy and modal is similar in Chong and Mazatec, but the magnitude of the contrast is not the same. Both languages have some 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 H1-H2 difference in Chong is never more than 3 dB, however, whereas the maximum difference in Mazatec is over 6 dB. This may be an actual linguistic phenomenon or it may be due to differences in recording levels.

Another important dissimilarity between the two languages is in the overall setting. The maximum H1-H2 in Chong breathy vowels is in the same range as the Mazatec modal vowels. Chong modal vowels are in the range of Mazatec laryngealized vowels. Thus both Chong phonation types may have a smaller open quotient than their Mazatec counterparts. The more laryngealized setting in Chong is plainly audible on the recordings.

4.2.1.2. H1-F2 (hypothesized indicator of abruptness of vocal fold closure)

Figure 23 shows the 8-speaker means for modal and breathy vowels. The two groups are significantly different for 200 ms, about 88% of the vowel (*p* values are given in table 20). This is unlike the Mazatec contrast, where breathy and modal vowels differ only for the first 50 ms, 33% of the vowel.

Table 20.
P *values of Chong H1-F2 differences*

ms	25	50	75	100	125	150	175	200	225
breathy vs. modal	.00006	.00002	.002	.04	.01	.01	.01	.001	.1

The dB values of both H1-H2 and H1-F2 are less in Chong than in Mazatec. The modal vowels at onset have an H1-F2 difference of 4 dB in Chong, 6 dB in Mazatec. The breathy vowels have an H1-F2 difference of about 12 dB in Chong, 16 dB in Mazatec. It is impossible to determine whether the smaller values in Chong are due to different recording conditions or are a fact of Chong phonetics.

4.2.1.3. Cepstral peak (periodicity)

Figure 24 compares the means of the cepstral peak prominences of breathy and modal vowels. The differences are significant except where the time courses cross at 75-100 ms, about 33% of the way through the vowel, (*p* values are given in table 21).

It is curious that at onset, breathy vowels are 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 (see figure 25); it is unlikely that pitch excursion is a factor.

Table 21.
P *values of Chong cepstral differences*

ms	25	50	75	100	125	150	175	200	225
breathy vs. modal	.01	.03	.6	1.0	.08	.04	.03	.11	.09

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.

4.2.1.4. Summary

Of the parameters studied, H1-F2 is the most important for differentiating Chong modal and breathy vowels. The amplitude of F2 relative to H1 is 5 to10 dB greater in breathy vowels, suggesting that the vocal folds close less abruptly. At vowel onset there is also a non-significant difference in H1-H2, suggesting that the open portion of the cycle is longer at the beginning of breathy vowels, but this parameter returns to modal values within about 50 ms, 22% of the vowel. Chong breathy vowels are characterized by a rapid increase in aspiration noise (diminution of cepstral peak) through the course of the vowel. In modal vowels the aspiration noise diminishes slightly at the vowel center, but does not change a great deal.

4.2.2. Mpi

The Mpi samples are unusually long in duration, probably because the speaker was pronouncing them carefully for the recording. We do not have recordings of connected speech to determine the normal duration of Mpi vowels.

Since the words all have [i] vowels, the influence of the first formant (F1) on the amplitude of H1 and H2 must be considered. A sampling of F1 frequency at 25, 150, 275, and 400 ms showed that F1 is at the same frequency as H2 in the modal vowels and the falling laryngealized vowels. Rising laryngealized vowels have F1 about 100 Hz higher than H2. Except at the very beginning and end, the relationship between F1 and H2 remains consistent throughout the vowel. The influence of F1 will be discussed further in the individual comparisons.

4.2.2.1. H1-H2 (hypothesized indicator of open quotient)

The H1-H2 means for modal and laryngealized vowels are shown in figure 26 for high falling tone, figure 27 for mid rising tone, and figure 28 for low rising tone. Note that the values for both phonation types are negative. This outcome is due to the proximity of H2 to the first formant in the [i] vowel. This speaker’s H2 ranges from 250 to 340 Hz, a frequency area that is amplified by the first formant (F1).

The first formant boosts the amplitude of H1 somewhat as well. If F1 boosts H2 more than H1, then H1-H2 will have a larger negative value. In the samples used here, the first formant was close to the frequency of H2 in high falling vowels of both phonation types. Thus the relationship shown between the two types in figure 26 is not affected by F1. But in the mid rising and low rising samples, the first formant was close to the frequency of H2 only in the <u>modal</u> vowels. Thus in figure 27 and figure 28 the values for H1-H2 are diminished for the modal vowels and not for the laryngealized vowels, putting the two lines closer together than they would be for [a].

For all tones the course of H1-H2 through the vowel is similar to that of F0. At high pitch the vowel may have a larger open quotient than at low pitch. It has not been determined whether this interaction is the result of an articulatory constraint. In the Mazatec sample, which was not matched for tone, the H1-H2 course of the high tone laryngealized vowel in [tSa^3] does not differ from that of the mid tone laryngealized vowels in [^{n}ka^2] and [Ba^2] (see figure 85.) This fact argues against the assumption of a human articulatory limitation as the cause for the H1-H2 pattern in Mpi. (The Mazatec 6-speaker male average F0 was 190 Hz for high and 180 for mid tone; the Mpi 1-speaker male average was 170 for high and 135 for mid tone.)

P values for all tones are given in table 22. Between the low rising groups there is no significant difference in H1-H2. The result may be due to the influence of F1. In the other groups, laryngealized vowels have a significantly smaller H1-H2 than modal vowels, starting at the 100th ms (24% of the way through the vowel) for high falling and the 75th ms (14% of the way through the vowel) for mid rising. The difference lasts for 225 ms (60% of the duration) in the high falling vowels and 175 ms (33% of the vowel duration) in the mid rising vowels. By contrast, Mazatec laryngealized vowels have a smaller H1-H2 than modal vowels for the entire course of the vowel.

For one comparison, inclusion of the /si/ words changed the value of H1-H2. In the modal high falling vowels, [si] had a value of H1-H2 that was 5 dB lower than that of [mi] for the first 300 ms. The difference is shown in figure 29. Omission of the [si] set would have resulted in a greater difference between modal and laryngealized vowels in figure 26 but would not have changed the essential finding.

4.2.2.2. H1-F2 (hypothesized indicator of abruptness of vocal fold closure)

On this measure the vowels in [si] had uniformly lower values than vowels in the other words, regardless of tone or phonation type. Since there were the same number of [si] samples in each tone and phonation type, their inclusion had no effect on the comparison <u>between</u> phonation types.

Table 22.
P *values of Mpi H1-H2 differences*

ms	25	50	75	100	125	150	175	200	225	250	275	300	325	350	375	400	425	450	475	500	525
low rising	.2	.2	.5	.6	.9	.4	.3	.8	.2	.3	.6	.7	.3	.6	.5	.4	.1	.06	.9		
mid rising	.1	.3	.02	.002	.03	.2	.05	.03	.06	.2	.3	.3	.2	.3	.1	.9	.4	.04	.5	.4	.7
high falling	.6	.8	.1	.04	.01	.01	.01	.01	.01	.02	.02	.05	.8	.3	.2	.02	.4				

The H1-F2 means for modal and laryngealized vowels are shown in figure 30 through figure 32. Assuming that H1 is boosted somewhat when F1 is at a low frequency due to the [i] vowel, the influence of F1 would augment a positive H1-F2 and diminish a negative H1-F2. Since F1 has the same frequency as H2 in high falling vowels of both phonation types, the relationship between the two types in figure 30 is not affected by F1. In the mid rising and low rising samples, F1 has the same frequency as H2 only in the modal vowels. Thus H1-F2 will be more positive for modal vowels, and the lines for the two phonation types in figure 31 and figure 32 may be farther apart than they would have been for [a].

The course of H1-F2, like that of H1-H2, parallels the fundamental frequency. At high pitch the vowel has a larger spectral slope than at low pitch. Thus on both measures, high F0 patterns with a more breathy setting. In the Mazatec sample, the H1-F2 course of the high and mid tone laryngealized vowels is essentially the same, once again providing evidence against an articulatory constraint as the reason for the H1-F2 pattern in Mpi.

P values for all tones are given in table 23. For high falling vowels, the difference in H1-F2 lasts only for the first 275 ms, 61% of the vowel, and is not always significant. In the other tone groups, the H1-F2 difference is robust (perhaps due to the influence of F1) and lasts for the entire course of the vowel. Although it appears that high falling vowels contrast primarily in H1-H2, while the other vowels contrast in H1-F2, this appearance may be due entirely to the influence of F1.

In rising-tone vowels, the spectral slope contrast is very similar to that of Mazatec, where there is a stable H1-F2 difference between the two phonation types throughout the course of the vowel. In Mazatec, the difference is about 10 dB; in Mpi, 5 dB. (The higher Mazatec value may be a linguistic fact or an artifact of the recording conditions.) But whereas in Mazatec, H1-H2 and H1-F2 work together to make the distinction, in Mpi they act independently: the distinguishing parameter depends on the tone. The interaction between tone and phonation parameter in Mpi may well have a physiological origin.

Table 23.
P *values of Mpi H1-F2 differences*

ms	25	50	75	100	125	150	175	200	225	250	275	300	325	350	375	400	425	450	475	500	525
low rising	.004	.02	.1	.07	.07	.1	.01	.005	.03	.04	.0002	.001	.1	.04	.01	.01	.002	.0006	.02	.1	.3
mid rising	.0008	.01	.003	.006	.0007	.01	.0004	.0003	.004	.0004	.001	.007	.0001	.003	.002	.05	.4	.02	.001	.1	
high falling	.04	.03	.09	.08	.2	.4	.2	.4	.1	.02	.01	.4	.2	.9	.7	.9	.6	.3			

4.2.2.3. Cepstral peak (periodicity)

On this measure, values for the vowels in [si] were not significantly different from vowels in the other words, regardless of tone or phonation type. I therefore combined the [si] words with the other words in this analysis.

The mean cepstral peak prominence for modal and laryngealized vowels is shown in figure 33 for high falling tone, figure 34 for mid rising tone, and figure 35 for low rising tone. In most samples the ratio of periodic to aperiodic sound increases for about 200 ms, then declines for the remainder of the vowel. In low rising vowels, the peaks range from 3 1/2 to 6 dB above the mean cepstral value, similar to the range in Mazatec and Tagalog. Peaks of mid rising vowels range from 5.5 to 8 dB above the mean, a higher overall periodicity than in Mazatec or Tagalog. High falling vowels traverse a wide range of cepstral values, rising from about 4.5 dB above the mean to a peak periodicity of 9 dB above the mean. Although *t* tests revealed no significant differences between phonation types, it appears that Mpi laryngealized vowels have a tendency toward more aperiodic sound than do modal vowels. P values are given in table 24.

Table 24.
P *values of Mpi cepstral differences*

ms	25	50	75	100	125	150	175	200	225	250	275	300	325	350	375	400	425	450	475	500	525	550	575
low rising	.8	.5	.9	.7	.3	.9	.6	.3	.3	.9	.3	.3	.2	.4	.8	.4	.3	.1	.06	.4	.5		
mid rising	.9	.8	.4	.8	.4	1.0	.5	.6	.9	.8	.9	.7	.7	1.0	.4	.8	.1	.8	.8	.7	.2	1.0	.5
high falling	.9	.6	.3	.3	.7	.2	.01	.09	.1	.1	.009	.001	.02	.02	.1	.2	.0006	.02					

4.2.2.4. Summary

The Mpi data set has only one speaker and nine samples of each type, which leads to poor statistical discrimination. Nonetheless, certain trends are apparent.

The most important parameter differentiating modal and laryngealized phonation is spectral slope (H1-F2), suggesting that the vocal folds close more abruptly in laryngealized vowels. In non-high vowels the difference is significant throughout the vowel. The amplitude of F2 relative to H1 is about 5 dB greater for modal low-rising vowels and 6 dB greater for modal mid-rising vowels than for laryngealized vowels of the same tone. It is impossible to determine how much of this difference is due to the influence of F1. H1-F2 differences of the same order of magnitude also appear during parts of the high falling vowels, whose comparison is not affected by F1.

The high vowels are also strongly differentiated by H1-H2, suggesting a longer open quotient in the breathy vowels, whereas these differences are weaker — or the measurement is neutralized by F1 — in the non-high vowels.

Periodicity may be slightly diminished in the laryngealized vowels, but does not appear to be a major factor in making the distinction between laryngealized and modal.

4.2.3. Navajo

The Navajo word list was not ideal for purposes of this experiment, since the target vowels in the glottal and h groups are followed by nasals or voiceless fricatives. Nasal zeros change the shape of the spectrum, altering the amplitudes of H1, H2, and F2, along with other landmarks. Voiceless fricatives require the vocal folds to separate, which can make the vowel more breathy. The glottal stop in [jah/anijood] is preceded by an [h]. Since the results for this word were not significantly different from those of the other words in the glottal group, the influence of [h] is minimal.

4.2.3.1. H1-H2 (hypothesized indicator of open quotient)

Figure 36 shows the 11-speaker means for the stop group and h group. Error bars indicate the standard error of the mean. *T* tests at each time interval show the groups to be significantly different only later in the vowel, (*p* values are given in table 25), and to become more different as they approach the following consonants. The vowels in the breathy group are followed either by a voiceless fricative, which induces spreading of the glottis during the latter part of the vowel, or by a nasal, which produces a zero near the frequencies of the lower formants. These two factors conspire to make the measurements increasingly “breathy” as the vowel progresses. At the first time interval, however, there is no significant difference in H1-H2, indicating that whatever influence the preceding [h] had on this acoustic factor had already ceased by the 25th ms.

Figure 37 shows the means for the stop group and glottal group. At the beginning of the vowel there is no significant difference between the groups, (*p* values are given in table 25), indicating that the influence of the [/], like that of the [h] in figure 36, was very short lived. It is striking that measurements for the glottal group also appear to become more “breathy” as the vowel progresses; once again the effect is due to the following nasal or voiceless fricative.

Table 25.
P *values of Navajo H1-H2 differences*

ms	25	50	75	100
[h] group vs. stop group	.14	.008	.0001	
[/] group vs. stop group	.4	.2	.2	.2

4.2.3.2. H1-F2 (hypothesized indicator of abruptness of vocal fold closure)

Figure 38 shows the 11-speaker means for the stop group and the h group; figure 39 shows the means for stop group and glottal group. *P* values for both comparisons are given in table 26. Once again there is no significant difference between the stop group and the other groups at the 25th ms. In all three groups, the trajectories of the H1-F2 means are very similar to those for H1-H2. In Navajo the articulations that affect H1-H2 and H1-F2 do not operate independently as in Chong and Mpi.

Table 26.
P *values of Navajo H1-F2 differences*

ms	25	50	75	100
[h] group vs. stop group	.3	.00009	.005	
[/] group vs. stop group	.6	.0002	.002	

4.2.3.3. Cepstral peak (periodicity)

Graphs comparing the time course of the cepstral peak prominence of the stop group to each of the other groups are shown in figure 40 and figure 41. *P* values for both comparisons are given in table 27. There are no significant differences between the stop group and either of the other groups. Thus once again, no influence of a preceding [h] and [/] can be observed, even as early as the 25th ms.

Table 27.
P *values of Navajo cepstral differences*

ms	25	50	75	100
[h] group vs. stop group	.2	.9	.7	.3
[/] group vs. stop group	.5	.3	.6	

4.2.3.4. Summary

None of the measures shows a consistent influence of [h] or [/] on the following vowel in Navajo. The vocal folds return to their canonical position before the 25 ms measurement, for which the window begins at 12.5 ms. The swift return to modal phonation shows that articulators can adjust rapidly to phonation requirements, and that modal phonation may be phonologically specified on Navajo vowels.

4.3. Summary of second experiment results

4.3.1. Breathy vowels

Table 28 summarizes the durations over which an expected characteristic occurred significantly in contrastively breathy vowels (Mazatec, Chong) or vowels following [h] (Tagalog, Navajo). Durations are given both in ms and as a percentage of the entire vowel. The contrast is at the beginning of the vowel unless otherwise noted.

All three parameters, H1-H2, H1-F2, and cepstrum, are exploited in Mazatec to achieve the contrast between breathy and modal vowels. Chong uses only H1-F2 and cepstrum. The duration of these properties may be controlled by the speaker, and the properties can vary separately. For example, in Mazatec breathy vowels, the breathy settings for both H1-H2 and H1-F2 are maintained for the same amount of time. But in Chong breathy vowels, the breathy value for H1-F2 is maintained separately, with H1-H2 at modal levels.

Among Navajo and Tagalog speakers, all three parameters had returned to their canonical vowel values before the first measurement at 25 ms. Thus the articulators can return to their normal position in less than 25 ms.

Table 28.
Durations of expected breathy values

CHARACTERISTIC	DURATION IN TAGALOG *allophonic*	DURATION IN NAVAJO *allophonic*	DURATION IN MAZATEC *phonemic*	DURATION IN CHONG *phonemic*
H1-H2 greater in breathy vowel (increased open quotient)	no effect	no effect	50 ms (33 %)	no effect
H1-F2 greater in breathy vowel (less abrupt vocal fold closure)	no effect	no effect	50 ms (33 %)	200 ms (88 %)
Cepstral peak amplitude less in breathy vowel (increased non-harmonic sound)	no effect	no effect	50 ms (33 %)	last 125 ms (55 %)

4.3.2. Laryngealized vowels

Table 29 summarizes the durations over which an expected characteristic occurred in contrastively laryngealized vowels (Mazatec, Mpi) or vowels following [/] (Tagalog, Navajo). Mazatec speakers used all three parameters to differentiate modal from laryngealized phonation. The Mpi speaker used only the gestures measured by H1-H2 and H1-F2, (along with a tongue root gesture not measured in this study.) Since the duration of these two parameters varied with tone, one can postulate that F0 frequency places limitations on their articulation. The tension associated with higher pitch, for example, may facilitate control of the open quotient but hinder control of closure abruptness.

Once again the Navajo results show that the articulators can return to their normal position in less than 25 ms. Thus the effects of the glottal stop on H1-F2 and cepstral peak prominence in Tagalog were not physiological requirements, but characteristics of the language. Tagalog speakers enhance the salience of the glottal stop by allowing its influence to last longer into the following vowel.

Table 29.
Durations of expected laryngeal values

CHARACTERISTIC	DURATION IN TAGALOG *allophonic*	DURATION IN NAVAJO *allophonic*	DURATION IN MAZATEC *phonemic*	DURATION IN MPI *phonemic*
H1-H2 less in laryngealized vowel (decreased open quotient)	no effect	no effect	125 ms (83 %)	High tone: 225 ms (60 %) Mid tone: 175 ms (33 %) Low tone: no effect
H1-F2 less in laryngealized vowel (more abrupt vocal fold closure)	25 ms (20 %)	no effect	125 ms (83 %)	High tone: 100 ms (22 %) Mid tone: 425 ms (85 %) Low tone: 475 ms (95%)
Cepstral peak amplitude less in laryngealized vowel (increased non-harmonic sound)	25 ms (20 %)	no effect	25 ms (17%)	no effect

5. Discussion

In the introduction, I posed three questions about the differences between languages where phonation type is contrastive and languages where it is not. This section will discuss how the results have answered the questions.

I. Is non-modal phonation of longer duration in languages with contrastive phonation type?

II. Is non-modal phonation more consistent across speakers in languages with contrastive phonation type?

III. Is non-modal 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 three additional topics in this discussion.

IV. Do all languages use the same acoustic cues?

V. Are there gender differences in how phonation contrasts are made?

VI. A cross-linguistic look at the laryngeal continuum

5.1. Duration

As has been shown in tables 28 and 29, the duration of non-modal phonation is consistently longer in the contrastive languages, both in absolute time and as a percentage of the complete vowel.

Nonetheless, the durations of breathiness in the contrastive languages Mazatec and Chong show no similarity, either in absolute duration or as a percentage of the whole vowel. Likewise there is no similarity in the durations of laryngeality in Mazatec and Mpi. There is apparently no optimum duration for a non-modal phonation type, so long as it lasts long enough to be perceptible.

The shortest durations of noncontrastive [h] and [/] effects on vowels in Tagalog and Navajo 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. Measuring peak airflow, Lofqvist and

McGowan (1992) observed durations of 52-58 ms for the effect of [h] and 57-75 ms for the effect of [/] on the following vowel in one Swedish and one English speaker. The Tagalog and Navajo data show that these longer durations in Swedish and English must be particular to the speaker or language, and not due to human articulatory limitations. Note also that the swift return to modal phonation in Tagalog and Navajo could indicate that modal is not simply a default position of the glottis for vowels, but is phonologically specified in those languages.

The shortest durations of contrastive phonation in Mazatec, Chong, and Mpi can instruct us about perceptual limits and about the interplay between articulatory effort and perceptual salience. In these data, it is rare for non-modal phonation to last through the entire vowel. There are several possible explanations for the result.

I. It may require extra effort to sustain a non-modal configuration of the vocal folds.

II. Surrounding segments may make conflicting articulatory demands, a possibility in the H1-H2 data for Chong breathy vowels, which are followed by voiceless stops.

III. There may be conflicting phonological 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 finding that Mazatec breathy vowels are breathy only during their first half. But there is no articulatory requirement that the breathiness be of short duration: the Chong data show us that a breathy spectrum can be maintained through the whole 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 has a distinct phonation pattern. The perception of phonation type in the absence of pitch information would be adequate to hear the distinctions.

On the other hand, Mazatec laryngealized vowels have laryngeal characteristics through a larger portion of their duration than do Mpi laryngealized vowels. The tone information in the Mpi data set allows us to see a possible articulatory explanation for the relative brevity of Mazatec laryngealization: while the Mpi mid- and low-tone vowels have a laryngealized spectrum for their entire length, the high-tone vowels do not. Perhaps the longitudinal tension of high tone interferes with the adduction required of laryngealization, which would otherwise last for the entire vowel. Mazatec is also a tone language, but since the data were not matched for tone, no clear judgment can be made about the interaction of tone and laryngeality. The laryngealized vowels in the sample did not include any low tone examples; the mid and high tone examples, which had similar F0 frequencies, behaved similarly on all three measures.

In each instance where the duration of non-modal phonation is dissimilar across languages, the difference can be explained by an articulatory or perceptual constraint on the shorter item. Thus we can posit that the default duration for nonmodal phonation is 100% of the vowel, and all variance from the default is in response to such constraints. If that is the case, then explanation 1 above — that extra effort is required to sustain nonmodal phonation — can be discarded. What remains is the familiar conflict between articulation and perception.

The duration data presented here for the Mazatec laryngealized vowels disagree with those of Silverman (1995). The difference is due to the method of measurement. Silverman's data are based on examination of spectrograms for an actual "creak', based on the earlier term "creaky vowel" for these items. There are frequent instances of laryngealized vowels in Mazatec where no creak is visible on the spectrogram or audible to the American ear. The spectrogram of such an instance is shown in figure 44, the Mazatec word [^{n}ka^{2}]. Compare this with spectrograms of a contrasting modal vowel in figure 42, [ka^{2}], and breathy vowel in figure 43, [^{n}da:^{23}], spoken by the same person.

Since native Mazatec speakers hear the creaky and noncreaky laryngealized vowels as members of the same type, they must be attending to another contrast. As can be seen in the spectra, both H1-H2 and H1-F2 give consistent reliable differences between the two phonation types. Thus a reduced H1-H2 and a more level spectrum are probable acoustic parameters for laryngealization. Creakiness when it occurs is a side effect of laryngeal tension. Creakiness is visible in the spectrogram of figure 45 [Ba^{2}], produced by the same speaker. The presence of laryngealization through the entire vowel is not a problem for Silverman's theory, since laryngealization does not interfere with tone perception as breathiness does. Breathiness and pitch are optimally perceptible if they occur on different parts of the vowel; laryngealization and pitch can be perceived just as well when they occur simultaneously.

5.2. Consistency

Figures 46 through 54 show the standard deviation for each measurement, summarized by language. Lower values indicate less variability across the speakers. The Mpi data are not included here because there was only one speaker. Certain measures — H1-H2 and H1-F2 for laryngealized vowels, and H1-H2 at the onset of breathy vowels — show less variation in the contrastive languages. But the cepstral peak for laryngealized vowels was more variable in the contrastive language. There is no consistent pattern of contrastive languages being less variable than noncontrastive languages. This result supports the notion that coarticulatory patterns may be specified in a language, in addition to the contrastive distinctions.

5.3. Magnitude of difference

Figures 55 through 60 show the magnitude of the differences between modal and nonmodal phonation for the measures used in this study, summarized across languages. Solid lines indicate languages where the difference is contrastive; dashed lines are languages where the difference is noncontrastive. Since no breathy samples were measured in Mpi, it does not appear in the figures that relate to breathiness; likewise Chong does not appear in the figures relating to laryngealization.

In figures 55, 56, and 57, I subtracted the average for modal vowels from the average for breathy vowels in the same time frame and language. For H1-H2 and H1-F2, breathy vowels usually have higher values than modal vowels; thus we expect the breathy minus modal difference to be positive. For cepstral peak, breathy vowels usually have lower values than modal vowels, giving a negative result for breathy minus modal. Since identical values for modal and nonmodal would give a result of zero, points farther from the zero line in either direction reflect greater differences between phonation types.

In figures 58, 59, and 60, I subtracted the average for modal vowels from the value for laryngealized vowels, usually producing negative values on H1-H2 and H1-F2, and positive values on cepstral peak.

Overall, the Tagalog vowels show smaller differences and the Mazatec vowels greater differences than those of the other languages. Since Tagalog and Mazatec were the languages of the main experiment, where the data sets could be more carefully controlled, this result may indicate a tendency for contrastive languages to have greater differences between phonation types. Chong and Mpi, the other contrastive languages, also show greater differences than Tagalog (except in the cepstral peak measurement for laryngealized vowels). But Navajo, which as a noncontrastive language should have smaller differences than Chong, Mpi, and Mazatec, often has larger differences instead. Therefore differences in contrastive languages are not always larger than in noncontrastive languages.

5.4. Does H1-H2 measure the same things as H1-F2?

Except for Tagalog, the shape of the curve for an individual language is similar between figure 55 and figure 56, and between figure 59 and figure 60. For example, the Mazatec breathy-modal difference is initially positive for both H1-H2 and H1-F2, decreases to zero between 75 and 100 ms, reaches a negative maximum at 125 ms, and then rises toward zero again. The Chong breathy-modal difference remains positive and fairly level on both measures throughout the vowel. The Navajo breathy-modal difference begins as a small positive value and increases; paradoxically the same pattern occurs for the laryngealized-modal difference in Navajo (figure 58 and figure 59), probably due to the effects of following consonants. (The Navajo word list was not well suited to this kind of study.) The

Mpi laryngealized-modal difference remains negative and level throughout the vowel. The Mazatec difference is even more negative, but approaches zero at the end of the vowel. The consistency between H1-H2 and H1-F2 gives the impression that the two measures record the same articulatory characteristic. But in Tagalog H1-H2 does not correlate with the spectral slope measurement H1-F2. Hanson (1997, 477) found that H1-H2 does not correlate with her spectral slope measurements (H1-F1 and H1-F3) for female English speakers. Thus H1-H2 and spectral slope apparently reflect separate articulatory parameters, although they are used together by some speakers.

5.5. The primary acoustic cue

With the variety of acoustic cues to phonation type, it would not be surprising to discover that different languages would choose different cues. There is some experimental evidence to show that this is the case.

In perceptual tests, Klatt and 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 *et al.* (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 6 native Gujarati speakers listening to synthesized Gujarati vowels modeled on natural speech. Instead, breathiness ratings correlated with increased H1-F2. The results of Ladefoged and Antonanzas-Barroso (1985) agree with those of Bickley and contradict those of Hillenbrand and Klatt & Klatt. Breathiness judgments by 10 American speakers listening to breathy and modal vowels by 10 !Xoo speakers were more strongly correlated with the spectral slope measure H1-F1 than with aspiration noise.

It is plausible that different languages prefer different cues for breathiness. The feature “breathy” encompasses a suite of articulatory gestures, among them reduced vocal fold closure duration (possibly resulting in a larger H1-H2 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, breathiness is instantiated with aspiration noise, in another with increased open quotient. Similarly, for laryngealization a language could employ changes in spectral slope, H1-H2, 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 members of the articulatory suite are used most in each language. On average, all humans are equally adept at making the possible articulations: differences between languages are not due to articulatory limitations. Thus if a language tends to favor one member of the suite of gestures, the preference may be based on acoustic or perceptual requirements. The preferred gesture for making the breathiness vs. modal contrast in Chong is one that changes the spectrum (more sloped for breathy than for modal), whereas Mazatec uses all the gestures about equally. The preferred gesture for making the laryngealized vs. modal contrast in Mpi is also one that changes the spectral slope (more level for laryngealized than for modal), while Mazatec uses changes in both spectral slope and H1-H2. It should be mentioned 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.

5.6. Gender differences

Recent studies (Henton & Bladon 1985, Klatt & Klatt 1990, Todaka 1993, Hanson 1997) have observed that female voices are breathier than male voices in English and Japanese. The breathier female setting is achieved by holding the arytenoid cartilages separated to form a chink where air can escape throughout the glottal cycle, by non-simultaneous or incomplete closure along the length of the glottis, or by a combination of these gestures. Opinions differ on whether the breathier female setting is due to physiological differences or cultural influences. Information from additional languages would help answer this question.

Although this study was not designed to determine gender differences, gender comparisons are facilitated by the fact that three of the language samples had equal numbers of male and female speakers. I shall present the gender data in order of increasing complexity, first Chong, then Tagalog, and finally Mazatec. Within languages, I will discuss each measure separately for all phonation types. Throughout this section, statistical differences are based on two-tailed *t* tests and a significance level of 0.05. As before, *P* values are shown in tables; cells with significant values are shaded.

5.6.1. Chong

The Chong gender breakdown for H1-H2 is presented in figure 61. *P* values of the gender differences are given in table 30. Within gender we see the expected relationship between phonation types. Across genders, female vowels are breathier than male vowels in each phonation type. The difference is significant on breathy vowels through 175 ms, 100% of the vowel, and on modal vowels through 150 ms, 57% of the vowel.

Table 30.
P *values of gender differences in Chong H1-H2*

ms	25	50	75	100	125	150	175
breathy	.02	.0004	.0003	.002	.01	.02	.01
modal	.007	.0009	.004	.003	.1	.04	.09

The breakdown for H1-F2 (figure 62) gives similar results. Within gender, breathy vowels have a greater H1-F2 (a more sloped spectrum) than modal vowels. Across genders, female vowels are significantly different from male vowels through 125 ms, 86% of the vowel, on breathy vowels, and 100 ms, 57% of the vowel, in modal vowels. *P* values of the gender differences are given in table 31.

Table 31.
P *values of gender differences in Chong H1-F2*

ms	25	50	75	100	125	150	175
breathy	.04	.002	.04	.002	.01	.06	.2
modal	.02	.0004	.04	.02	.09	.04	.05

The breakdown for cepstral peak prominence (figure 63) gives an interesting picture. Each gender shows the curious pattern discussed previously for the combined genders: breathy vowels are more periodic than modal vowels at the outset. Between genders, female vowels are more periodic than male vowels in both phonation types. Although the difference is not significant on modal vowels, and is significant only after 125 ms on breathy vowels, it appears that reduced periodicity is not a distinguishing characteristic of Chong female speech. *P* values of the gender differences are given in table 32.

Table 32.
P *values of gender differences in Chong cepstral peaks*

ms	25	50	75	100	125	150	175	200
breathy	.7	.3	.8	.3	.08	.009	.04	.04
modal	.5	.8	.2	.2	.2	.6	.9	.13

5.6.2. Tagalog

As in earlier sections, I shall refer to vowels following [h] as the “h group”, vowels after [/] as the “glottal group”, and vowels after non-glottal stops as the “stop group.” The Tagalog gender breakdowns for H1-H2 are presented in figure 64 for the h and stop groups, and figure 65 for the glottal and stop groups. *P* values of the gender differences are given in table 33. Female vowels are breathier than male vowels in the h group and stop group, but there is no difference between genders in the glottal group. Differences in the h group are significant at 25-75 ms, 60% of the vowel, and in the stop group at 50-100 ms, 60 % of the vowel. Within gender, the female vowels show the expected relationship between phonation groups (h group > stop group > glottal group) for the first half of the vowel, but the male vowels after [/] lie in a more “breathy” range than the other two phonation groups.

Table 33.
P *values of gender differences in Tagalog H1-H2*

ms	25	50	75	100	125
[h] group	.02	.009	.04	.1	.3
stop group	.06	.03	.005	.03	.5
[/] group	.7	.9	.9	.5	.3

The breakdown for H1-F2 is shown in figure 66 for the h and stop groups, and figure 67 for glottal and stop groups. *P* values of gender differences are given in table 34. Female vowels are breathier than male vowels in all three groups. The differences are significant at 25 ms, 20% of the vowel, for the h and stop groups, and 25-100 ms, 80% of the vowel, for the glottal group. For both genders, vowels after stops begin with a higher measure (a more sloped spectrum) than those after [h], but switch to a lower measure by 50 ms. In the female group (figure 67), the spectrum-flattening effect of [/] lasts only until the 50th ms.

Table 34.
P *values of gender differences in Tagalog H1-F2*

ms	25	50	75	100	125
[h] group	.02	.1	.2	.3	.4
stop group	.03	.08	.2	1.0	1.0
[/] group	.03	.003	.004	.03	.1

The gender breakdown for cepstral peak prominence (figure 68 and figure 69) is similar to the one for Chong. *P* values of gender differences are given in table 35. On the whole, female vowels are more periodic than male vowels. Reduced periodicity is thus not a characteristic of female vowels in Tagalog. The differences are significant at 50 and 100-125 ms in the h and stop groups, and not significant in the glottal group.

Table 35.
P *values of gender differences in Tagalog cepstral peaks*

ms	25	50	75	100	125
[h] group	.1	.01	.2	.05	.05
stop group	.3	.03	.07	.003	.04
[/] group	.8	.4	.6	.05	.4

In data presented earlier for the combined genders, vowels in the glottal group have a lower peak than vowels in the stop group. Vowels in the h group and stop group have the same cepstral peak prominence, except that the h group is slightly higher at the outset. The same relationships hold true within the separate genders.

5.6.3. Mazatec

The Mazatec gender breakdown for H1-H2 is presented in figure 70. *P* values are given in table 36. Female modal vowels are significantly breathier than male vowels at 50-125 ms, about 80% of the vowel. There is no significant difference between genders in the other phonation types. Within gender we see the expected relationships between phonation types.

Table 36.
P *values of gender differences in Mazatec H1-H2*

ms	25	50	75	100	125	150	175
breathy	.06	.07	.7	.6	.5	.5	.8
modal	.3	.04	.05	.003	.03		
laryngealized	.3	.09	.2	.2	.3	.07	

In the breakdown for H1-F2 (figure 71), the within-gender relationships are in the expected order, but there is no significant difference between genders in the breathy and modal types. *P* values are given in table 37. The significant difference at 25-75 ms on laryngealized vowels is in the opposite direction to that expected: the male vowels are more breathy (have a more sloped spectrum) than the female.

Table 37.
P *values of gender differences in Mazatec H1-F2*

ms	25	50	75	100	125	150	175
breathy	.8	.1	.3	.06	.08	.08	.06
modal	.1	.8	.4	.8	.8		
laryngealized	.04	.01	.006	.2	.7	.3	

Figure 72 and figure 73 show the gender breakdown for cepstral peak prominence. *P* values are given in table 38. Within gender the relationships are in the expected order at the outset, with higher peaks (more periodic sound) for modal than the other phonation types. Between genders there are no significant differences of interest, except that at the 25th ms of the modal phonation type, female vowels are *more* periodic than male vowels.

Table 38.
P *values of gender differences in Mazatec cepstral peaks*

ms	25	50	75	100	125	150	175
breathy	.2	.1	.8	1.0	.3	.7	.7
modal	.02	.2	.07	1.0	.2		
laryngealized	.07	.3	1.0	.7	.3	.03	

5.6.4. Summary

Table 39 summarizes the gender differences found in the three languages. The characteristics expected for a more breathy female setting are: greater H1-H2, greater H1-F2, and lower cepstral peak prominence. F indicates the average female values and M the average male values. A cell with less than (<) or greater than (>) but no asterisk indicates a non-significant but visible difference for at least half of the duration of the vowel. An asterisk indicates that significant differences existed for at least two time windows (50 ms). An equals sign (=) indicates that no difference was observed. Where results vary from one phonation type to another, the separate phonation types are indicated by *b* for breathy or h group, *m* for modal or stop group, and *l* for laryngealized or glottal group.

For the H1-H2 measure, all three languages yield the expected relationship on modal and breathy vowels, but there is no difference between genders on laryngealized vowels. In articulatory terms, the female voices may be more breathy by virtue of an increased open quotient, but this longer open quotient is not (and maybe cannot be) maintained in the presence of laryngealization. Henton & Bladon (1985), Klatt & Klatt (1990), and Hanson (personal communication) obtained similar results for this measure in English. Thus H1-H2 is quite a reliable indicator of gender differences.

For the H1-F2 measure, Chong and Tagalog yield the expected relationship in all phonation types, but Mazatec female vowels have *less* spectral slope than male vowels in nonmodal phonation, and there is no gender difference in modal phonation. Several explanations for this finding are possible:

I. Mazatec female vowels are not breathier.

II. Mazatec female vowels are breathier but only by means of a larger open quotient.

III. The measure H1-F2 does not work well for discriminating between genders.

Explanation 1 is disproved by the results for H1-H2 above. Explanation 2 cannot be proved or disproved with the current data. Explanation 3 receives some support from the disparity in results for English. Hanson's results were in the same direction (F>M) as the Chong and Tagalog results, but Klatt & Klatt's results were in the opposite direction (F<M), although not significantly. Since each study used a different approach to obtaining the amplitude of F2, the results may not be comparable across studies. The results for the three languages presented here are comparable, however, and Mazatec diverges from the others.

Table 39.
Summary of gender differences. (Asterisks mark significant differences.)

Measure	Expected relationship	Chong	Tagalog	Mazatec
H1-H2	F>M	F>M*	*b* F>M*	*b* F>M*
			m F>M*	*m* F>M*
			l F=M	*l* F=M
H1-F2	F>M	F>M*	*b* F>M	*b* F<M
			m F>M*	*m* F=M
			l F>M	*l* F<M*
Cepstral peak prominence	F<M	F>M	*b* F>M*	*b* F<M
			m F>M*	*m* F>M*
			l F=M	*l* F=M

For the cepstral peak measure, only Mazatec breathy vowels yield the expected relationship. In all three languages, female modal vowels were significantly more harmonic than male modal vowels. There is no difference between genders in laryngealized vowels.

Although other gender studies have not used cepstral peak prominence, they have used other measures related to periodicity. (Todaka (1993) employed jitter, shimmer, and harmonics to noise ratio of the electroglottograph signal; Klatt & Klatt (1990) and Hanson (p.c.) used noise judgments made on spectrograms). All of those studies found female speech to be less periodic than male. Although we cannot rule out the possibility that cepstral peak is an inappropriate measure, the consistent significant results for Chong, Tagalog, and the Mazatec modals suggest that reduced periodicity may not be a defining trait of female vowels, but may vary from language to language.

In summary, gender differences are not realized in the same way across languages, and may not exist in all languages. The most reliable discriminator of gender is H1-H2, and that difference may in fact exist in all languages. Studies of other languages will be required in order to clarify the question. If the H1-H2 gender difference is universal, then an articulatory and physiological basis can be posited. The lack of a gender distinction in laryngeal context would not be a problem for such a theory, since laryngealization requires a stiffening of the vocal folds that could be expected to interfere with the distinction.

It is interesting that the gender difference on H1-H2 in English and Chong is on the order of 5-6 dB, while in Tagalog it is about 4 dB, and in Mazatec 2-3 dB. The lack of consistency across languages may be due to different recording conditions, but it may indicate the presence of culturally based variation, even if the essential difference is physiological.

5.7. The laryngeal continuum

The body of research that commenced with Lisker & Abramson (1964) established that voice onset time (VOT), the time at which voicing begins relative to the release of a consonant, is a continuum of possibilities within which each language has from one to four contrastive categories. The locations of the categories 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 distinguishing between two items that are within the same category in their native language.

This study illuminates another such continuum, degree of laryngeal tension. (Laryngeal tension is a term of convenience. It is beyond the scope of this paper to determine whether the articulatory variation is based on longitudinal tension, ab-/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. It has recently 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, 345). Many languages have two categories of vowels on the glottal stricture continuum, among them 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. Pamela Munro (personal communication) is currently studying a variety of Zapotec that may have four phonation categories.

As with VOT categories, locations of categories on the laryngeal continuum vary from language to language. The phonation categories of the sample languages illustrate this point. Table 40 compares the male H1-H2 averages during the first half of the vowel in figure 70 (Mazatec), figure 61 (Chong), and figure 27 (Mpi). (These vowels are as nearly comparable as possible. Except for one of the three Mazatec words, all vowels begin on mid-tone. The Mpi vowel is [i], the others [a]. Number of speakers and number of samples are not the same.) Differences in H1-H2 may simply be due to differences in recording conditions, or they may indicate actual category locations on the phonation continuum. The three Mazatec categories all have positive values for H1-H2. Chong values are negative, and Mpi values even more negative, possibly because of the [i] vowel. 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 40.
Male H1-H2 averages (in dB) for each phonation type

PHONATION TYPE	MAZATEC	CHONG	MPI
breathy	9	-1	
modal	4	-3	-10
laryngealized	1		-12

Table 41 compares the male H1-F2 averages during the first half of the vowel. These form a similar pattern to the H1-H2 averages. This time the Chong and Mpi modals have the same value, but once again a Chong breathy vowel might be heard as a laryngealized vowel by a Mazatec speaker.

Table 41.
Male H1-F2 averages (in dB) for each phonation type

PHONATION TYPE	MAZATEC	CHONG	MPI
breathy	16	6	
modal	9	0	0
laryngealized	5		-6

5.8. Directions for further research

5.8.1. Perceptual cues

Many of the speculations in this paper have to do with perceptual cues. Although no perceptual measurements were available, areas of acoustic consistency point to possible perceptual cues in the spectrum and their time course through the vowel. It would be interesting to know if any of these is actually exploited by listeners, and whether the same cues are important in each of the languages. Little perceptual testing has been done in field work situations, but such measurements are now feasible and should be introduced wherever possible.

Stimuli for a perceptual test might be based on a potential minimal pair in the target language, with the vowels synthetically altered to create samples with a range of spectral slopes. Native listeners' identification of the samples as one word or the other would provide information on whether spectral slope is a salient parameter in the distinction and where the category boundary is located.

5.8.2. Consonants

This study has considered phonation only on vowels. The laryngealization continuum has an even larger range on consonants, from /?/ through glottalized, unmarked, and breathy consonants, to /h/. Since the effects of glottal tension on consonants spill over to the surrounding vowels, the methods developed in this study could be used to detect glottal tension on consonants. For instance, vowel phonation is an acoustic and perceptual cue to Korean initial plosives (Lisker & Abramson 1984, 397; Cho 1996). The aspirated and glottalized consonants of Mazatec and related languages offer other intriguing paths for further research.

6. Concluding summary

This research has examined three elements of nonmodal phonation: duration, magnitude, and inter-speaker variability. Measures of H1-H2, spectral slope, 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. Contrary to expectations, there was not less inter-speaker variability in languages with a phonation contrast.

The results showed that the vocal folds can return from /// or /h/ to modal phonation is less than 25 ms, but that it is possible to sustain nonmodal phonation through an entire vowel when it is required. The study provides further evidence that the gestures associated with H1-H2 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. Similarly, gender differences are not realized on the same cue across languages, and may not exist in all languages.

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 indicates that from the continuum of possible spectra, each language establishes a range of spectra for its phonation categories.

7. Appendix A

Single word averages from main experiment

Like the combined-word data in the main experiment, the results for individual words will be presented in three sections, H1-H2, H2-F2, and cepstral peak. Within each section, I will give the Tagalog data first, followed by the Mazatec data. The order of presentation for each language is as follows.

I. Individual words within the modal set: the time course of the 12-speaker averages for the measure under consideration, possible reasons for differences between words

II. Individual words within the breathy or after-h set

III. Individual words within the laryngealized or after-/ set

The figures use the same scales as those of the main experiment. Due to individual differences in vowel duration, time frames near the ends of vowels may have fewer than 12 speakers remaining. Time frames where fewer than 4 of the 12 samples remain are omitted from consideration.

7.1. H1-H2

7.1.1. Tagalog

The amplitude differences between H1 and H2 tend to be uniform across the entire duration of vowels in the stop group, as shown in figure 74.

Within the h group (see figure 75), [mahal] and [mahalai] show the influence of the preceding [h] by high values of H1-H2 at the outset, indicating that the glottis is open for a longer portion of the cycle. Toward the end of the syllable, all three vowels have a decreasing trend in H1-H2 preceding continuants.

Vowels of the glottal group (see figure 76) have a relatively level H1-H2 difference until the midpoint. The closure of the Tagalog glottal stop does not change the position of the vocal folds enough to have a large effect on the vocal-fold closure duration of the following vowel onset.

During the second half of the vowel, H1-H2 of the word-final vowel in [pa/a] increases, indicating that the glottis stays open for a longer portion of the cycle immediately preceding the utterance-final vowel release. In the second half of the vowels preceding continuants [pa/alam, taga/ajos], H1-H2 decreases slightly preceding continuants, as it does in the h group.

7.1.2. Mazatec

The H1-H2 means for modal vowels are shown in figure 77. The amplitude difference between H1 and H2 increases throughout the modal vowels, indicating that the open portion of the glottal cycle gradually grows longer. The data accord with the auditory impression that Mazatec words in isolation typically have a breathy ending. The low-tone word [na^1] has consistently higher values than the mid-tone words [ka^2, ^{n}da^2].

Like the modal vowels, the laryngealized vowels show a rising trend (figure 79) but H1-H2 means for the laryngealized vowels are consistently lower than those of the modal vowels (figure 12). The tenseness of the vocal folds in laryngealized phonation keeps them closed for a longer portion of the vibratory cycle.

7.2. Spectral slope: H1-F2

7.2.1. Tagalog

Vowels of the stop group are shown in figure 80; those of the h group are shown in figure 81. The two groups show a similar pattern. The amplitude difference between H1 and F2 ranges from 0 to 5 dB for most of the vowel, rising at the end. The rise indicates vocal cord relaxation at the end of the stressed syllable, allowing the cords to meet more gradually along their length.

Vowels of the glottal group (figure 82) display the same rising pattern at the end of the stressed syllable, but they begin at zero or slightly below: F2 has an equal or higher amplitude than H1 at the outset of the vowel due to the tension of the vocal folds at the release of the glottal stop.

It is interesting that H1-F2 stays low (the spectral slope remains relatively level) near the close of the only word-final vowel in the sample, [pa/a]. Figure 76 showed that the folds were moving further apart at the end of [pa/a]. Figure 82 shows that the folds are also somewhat stiffened, such that their meeting is more like a slap than an undulation.

7.2.2. Mazatec

The H1-F2 means for modal vowels are shown in figure 83. The amplitude difference between H1 and F2 increases throughout the modal vowels, indicating that the abruptness with which the vocal folds meet gradually declines throughout the vowel.

The H1-F2 means for breathy vowels are shown in figure 84. The mean H1-H2 difference is high during the first 50-75 ms, reaches a low point at 100-125 ms, then rises again. These results parallel those for H1-H2 (figure 78.)

Like the modal vowels, the laryngealized vowels show a rising trend (figure 85), but they have consistently lower values of H1-F2. The tenseness of the vocal folds in laryngealized phonation causes the vowel folds to slap together abruptly, which gives higher amplitudes throughout the frequency spectrum and a higher amplitude of F2 relative to H1. F2 is actually higher than H1 (H1-F2 is negative) at the onset of the laryngealized vowels in the sample.

7.3. Periodicity: cepstral peak prominence

7.3.1 Tagalog

Figures 86, 87, and 88 show the cepstral peak prominence for the Tagalog stop, h, and glottal groups. The peak ranges from 3 1/2 to 6 dB above the mean cepstral value. In most samples the ratio of periodic to aperiodic sound increases until the 75th ms, then declines rapidly. Vowels followed by [j] show an increase in periodicity at the end.

7.3.2. Mazatec

Figures 89, 90 and 91 show the cepstral peak prominence for the Mazatec modal, breathy, and laryngealized vowels. The peak ranges from 2 to 6 dB above the mean cepstral value, a wider range than that of Tagalog. Once again, the low tone word [na^{1}] has a more breathy setting than its mid-tone counterparts.

Among the modal and laryngealized samples, [tSa^{3}] is the only syllable that increases to its peak periodicity at the 75th ms like the Tagalog examples. The rest have the greatest ratio of periodic to aperiodic sound at the outset of the vowel. During the first 75-100 ms of the breathy vowels, however, there is a low ratio due to airflow noise through the glottis.

8. Appendix B

Single word averages from second experiment

This section gives single word data for Chong and Navajo. Mpi was not included, because the words had variable numbers of samples and there was only one speaker. Like the combined-word data in the main experiment, the results for each language will be presented in three sections, H1-H2, H2-F2, and cepstral peak. Within each section, the order of presentation is as follows.

I. Individual words within the modal set: the time course of the all-speaker averages for the measure under consideration, possible reasons for differences between words

II. Individual words within the breathy or after-h set

III. Individual words within the laryngealized or after-/ set

The figures use the same scales as those of the main experiment. Due to individual differences in vowel duration, time frames near the ends of vowels may have fewer than 12 speakers remaining. Time frames where fewer than one third of the samples remain are omitted from consideration.

8.1. Chong

8.1.1. H1-H2

The H1-H2 means for Chong modal vowels are shown in figure 92. The amplitude difference between H1 and H2 remains fairly level until the end of the vowel, where it increases, suggesting that the open portion of the glottal cycle grows longer at the end. Mazatec modal vowels, by contrast, have a steady increase in H1-H2 difference throughout the course of the vowel.

The H1-H2 means for Chong breathy vowels are shown in figure 93. The amplitude difference between H1 and H2 begins at about 2 dB, and decreases gradually over the first 125 ms of the vowel. Mazatec breathy vowels show the same H1-H2 pattern, but in a higher range of values.

8.1.2. H1-F2

The H1-F2 means for Chong modal vowels are shown in figure 94. The amplitude difference between H1 and F2 remains fairly level through the vowel, in contrast to Mazatec modal vowels, which show a steady increase in H1-F2 throughout the course of the vowel.

The H1-F2 means for Chong breathy vowels are shown in figure 95. The mean H1-F2 difference is high during the first 50-100 ms, reaches a low point at 125-150 ms, then rises again. These results parallel those for Mazatec.

8.1.3. Cepstral peak

Figures 96 and 97 show the time course of the cepstral peak prominence for the Chong modal and breathy vowels. In modal vowels, the periodicity increases for the first 125 ms, then decreases, within a narrow range of about 2 dB. Periodicity of the breathy vowels ranges over 4 dB, decreasing continuously from a peak that occurs within the first 50 ms.

8.2. Navajo

8.2.1. H1-H2

The H1-H2 means for stop group (a single word) are shown in figure 98. The amplitude difference between H1 and H2 remains level at about 4 dB throughout the vowel, just as in the Tagalog sample. In both languages the sample vowel is between voiced consonants.

The H1-H2 means for the [h] vowels are shown in figure 99. The more breathy overall setting of the vowel in [nahaS/na] may be due to its context between two voiceless fricatives that require abducted vocal folds. The amplitude difference between H1 and H2 for all three words increases by 3-4 dB over the first 125 ms of the vowel. Two vowels of the glottal group (see figure 100) show the same pattern. This is the only language in the sample where such an increase takes place in non-final vowels.

8.2.2. H1-F2

The H1-F2 means for stop group are shown in figure 101. The amplitude difference between H1 and F2 remains level through the vowel.

The H1-F2 means for the h group are shown in figure 102 and those for the glottal group in figure 103. In all three groups, the trajectories of the H1-F2 means are very similar to those for H1-H2 (figures 98-100, above). In Navajo the articulations that produce H1-H2 and H1-F2 do not operate independently as in the other sample languages.

8.2.3. Cepstral peak

Figure 104, figure 105, and figure 106 show the time course of the cepstral peak prominence for individual words of the three groups. The peak ranges from 2 to 5 dB above the mean cepstral value.

Figures

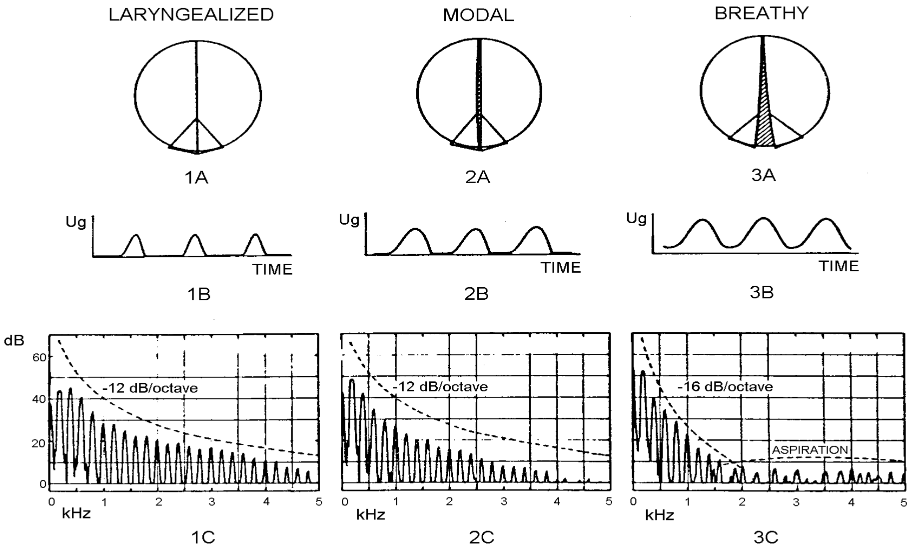


Figure 1.
Hypothetical glottal configurations of the phonation types,
from Klatt & Klatt (1990, 822).

Top row: schematic configurations for laryngealized (1A), modal (2A), and breathy (3A) vowels.

Middle row: waveform of the glottal vibration for each phonation type. Note the increase in open quotient from (1B) to (2B) and the lack of closure in (3B).

Bottom row: the resulting spectra. (3C) has additional aspiration noise due to incomplete glottal closure.

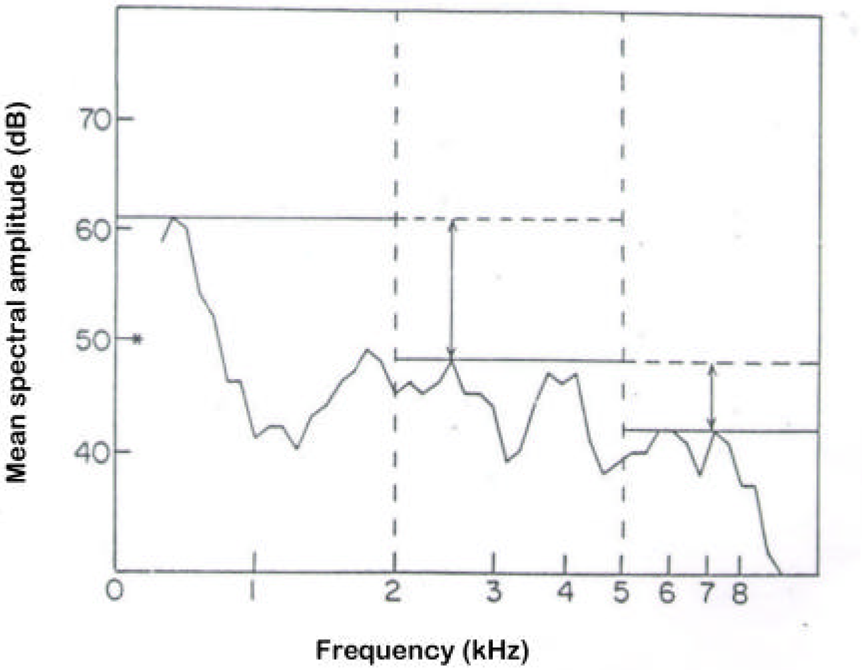


Figure 2.
Comparison of three frequency ranges of a breathy vowel spectrum, from Hammarberg & Gauffin (1995, 292).

The horizontal axis represents frequency in kHz; the vertical axis represents amplitude in dB. An asterisk indicates the amplitude of the fundamental frequency.

The frequency scale is subdivided into three ranges: 0-2 kHz, 2-5 kHz, and 5-10 kHz. The amplitude of the highest peak within each range is used to estimate the spectral slope.

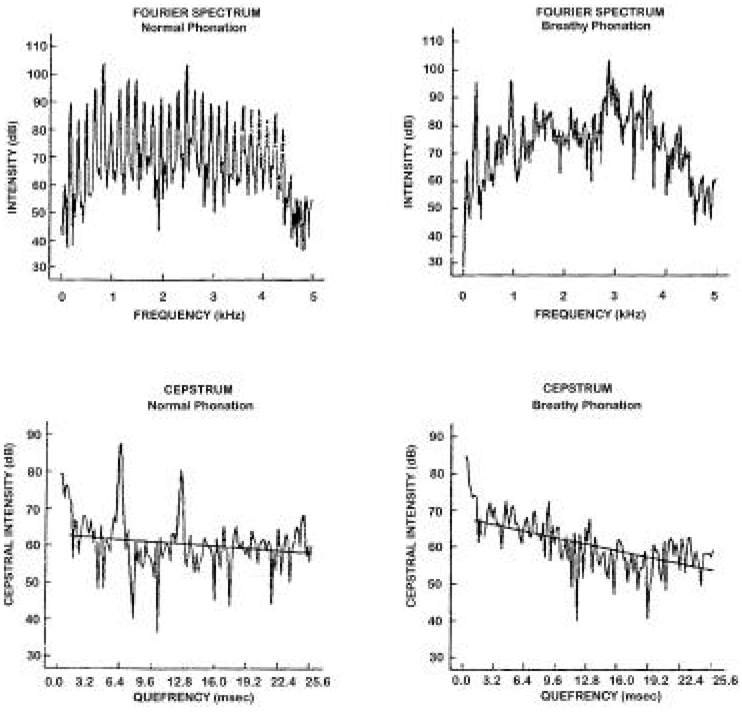


Figure 3.
Fast Fourier Transform (FFT) spectrum and cepstral representation of modal and breathy vowels, from Hillenbrand et al. *(1994, 772).*

The upper left illustration shows the spectrum of a modal vowel; the lower left illustration is the cepstrum of the same vowel. Illustrations on the right show the spectrum and cepstrum of a breathy vowel.

The linear regression line of the cepstra is used to normalize the peaks for overall signal amplitude. The location of the highest peak on the x-axis scale (the "quefrency") corresponds to the fundamental period of the signal.

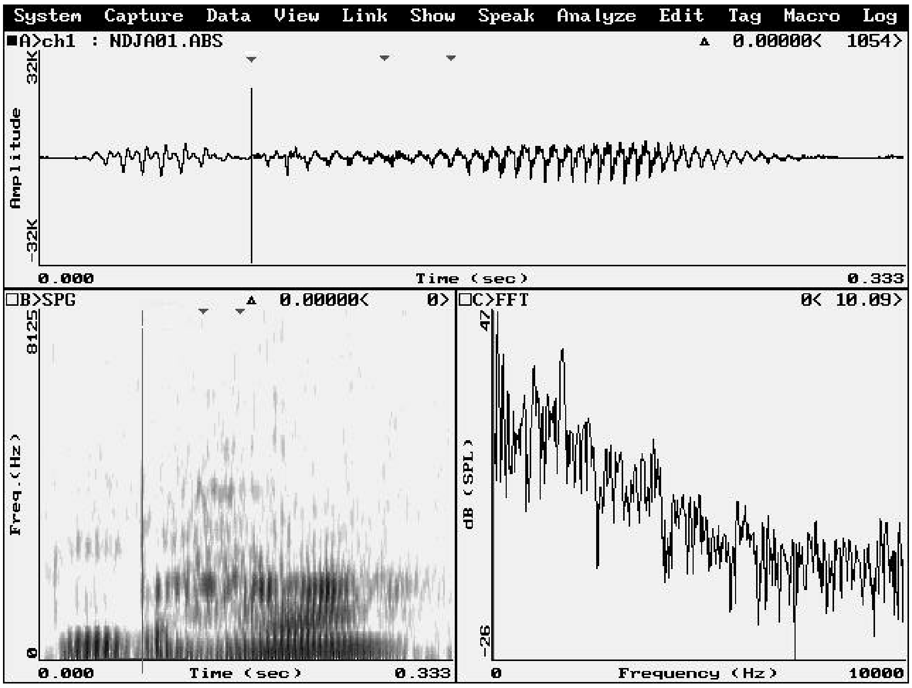


Figure 4.
Location of test window for methods trial.

Window A: sound wave. The cursor is at vowel onset. Other arrows are at the 50th and 75th millisecond.

Window B: spectrogram of the same sound wave. Cursor is at vowel onset. Other arrows are at the 50th and 75th millisecond. Note that the vowel is still breathy between the arrows, becoming modal later.

Window C: Fast Fourier transform (FFT) of the interval between the arrows. For the methods trial, the amplitudes of H1, H2, and the highest peaks in the first two formants were taken from this FFT spectrum.

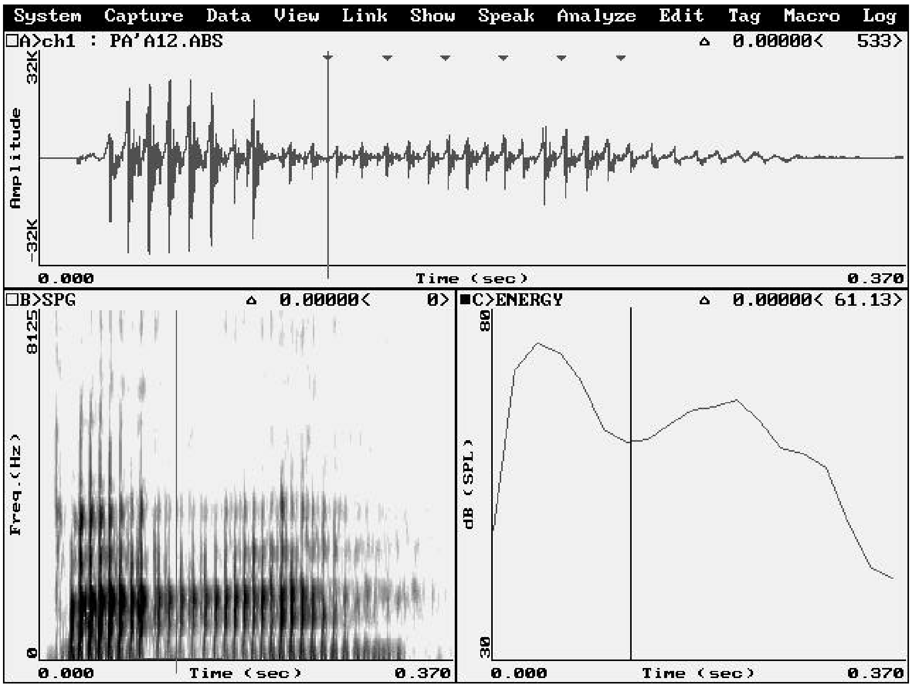


Figure 5.
Use of energy measure to determine vowel onset.

Window A: Waveform of the Tagalog word [pa/a] ‘foot’.

Window B: Spectrogram of the waveform.

Window C: Energy measure of the waveform.

Since onset of the second vowel cannot be determined from A or B, the point of minimum energy in C was taken as the onset. Cursors in each window show this point.

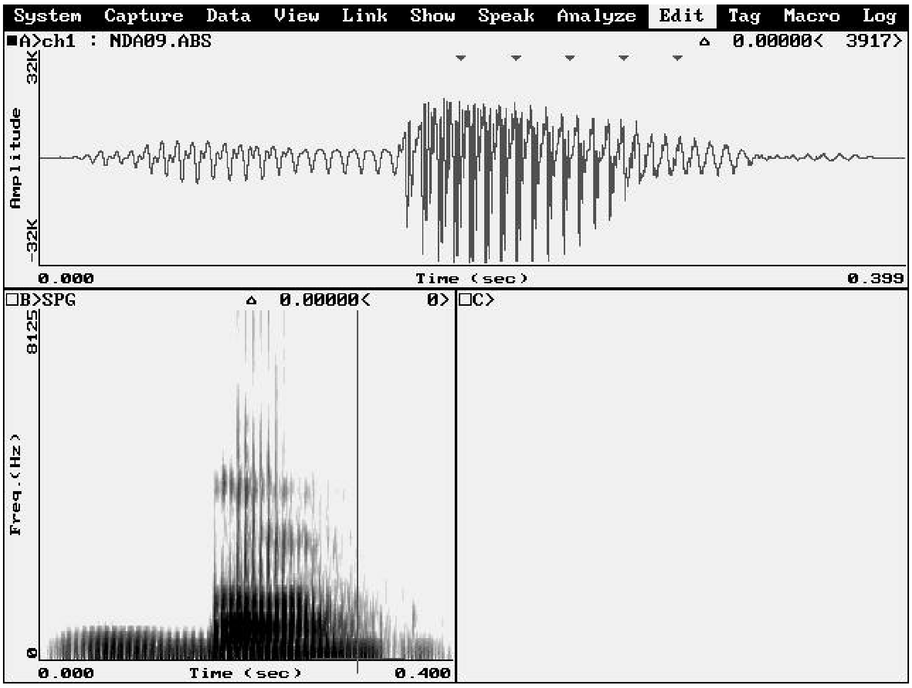


Figure 6.
Placement of tags for measurements.

Window A: Waveform of the Mazatec word [^{n}da^{2}] ‘good’.
Window B: Spectrogram of the waveform.

Tags are at 25 ms intervals from vowel onset. The cursor in Window B coincides with the final tag in Window A. Although voicing continues after the cursor, the formants do not. The portion after the cursor was not included in the analysis.

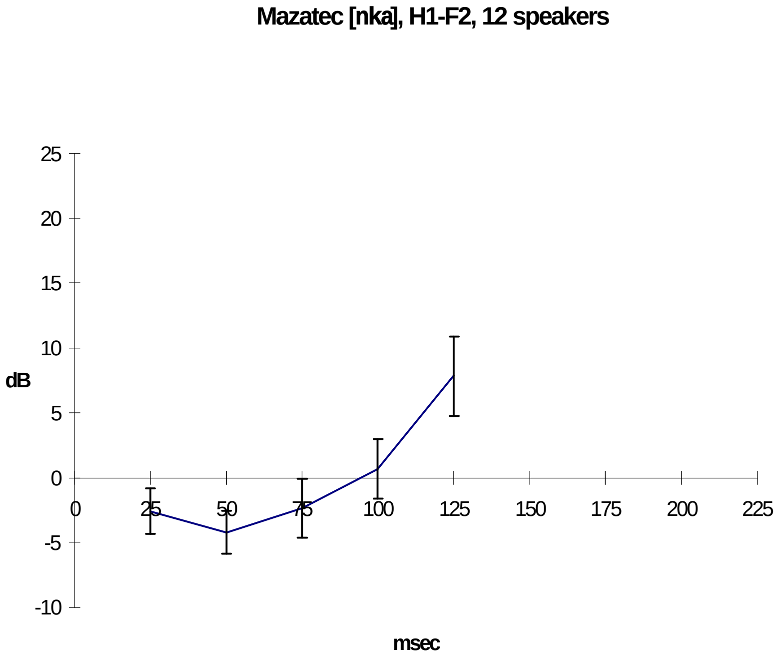


Figure 7.
Sample graph for one Mazatec word with the twelve speakers averaged.

Horizontal axis represents time from vowel onset, in milliseconds. Vertical axis represents amplitude in decibels. The figure shows the time course of H1-H2 through the vowel in [^{n}ka^2] ‘high’.

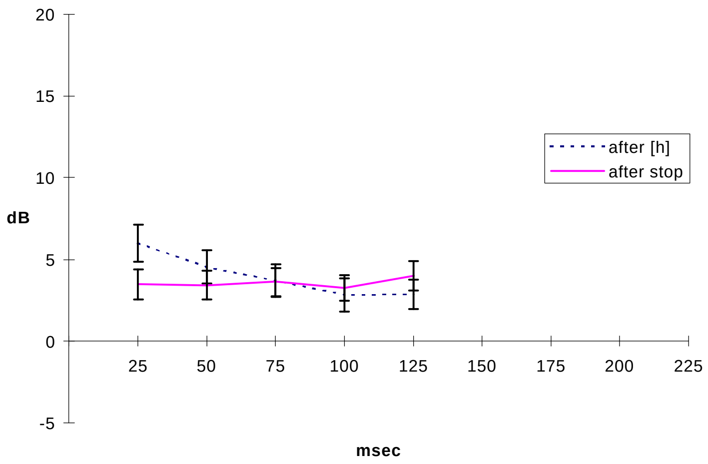


Figure 8.
TAGALOG, H1-H2, 12 speakers. Vowels after [h] vs. vowels after stop.

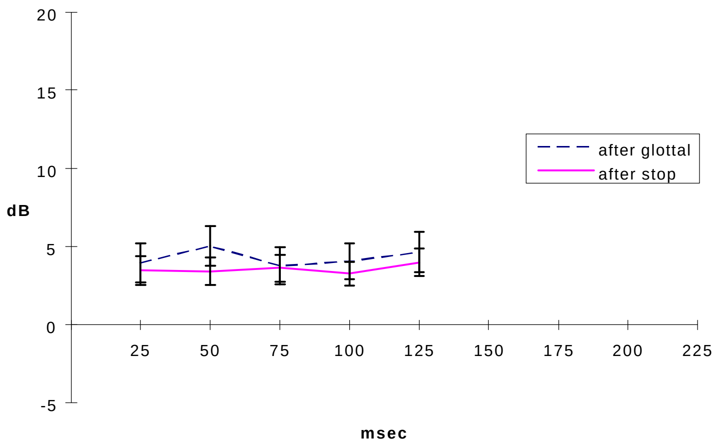


Figure 9.
TAGALOG, H1-H2, 12 speakers. Vowels after [/] vs. vowels after stop.

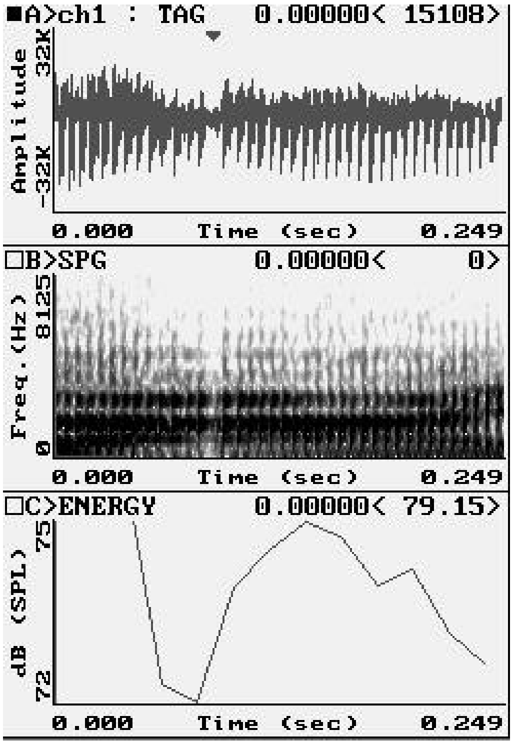


Figure 10.
[a/a] excerpted from the Tagalog [taga/ajos] 'manager', speaker 1, male.
The location of the glottal stop is marked by a triangular tag in Window A.
All three windows have the same time range.

Window A: Waveform
Window B: Spectrogram of the waveform.
Window C: Energy measure of the waveform.

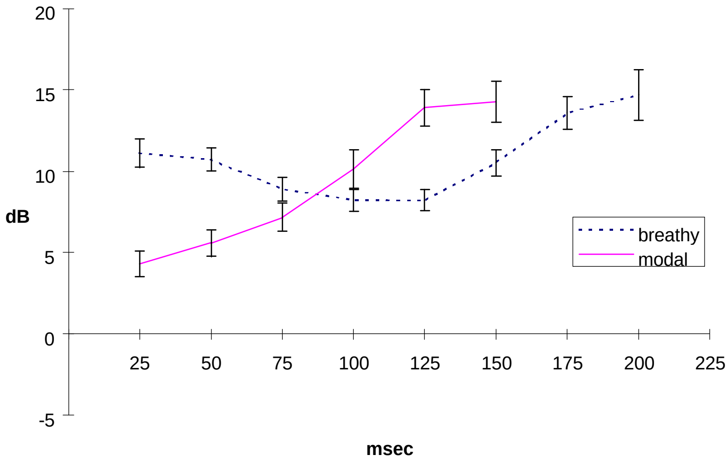


Figure 11.
MAZATEC, H1-H2, 12 speakers. Breathy vs. modal vowels.

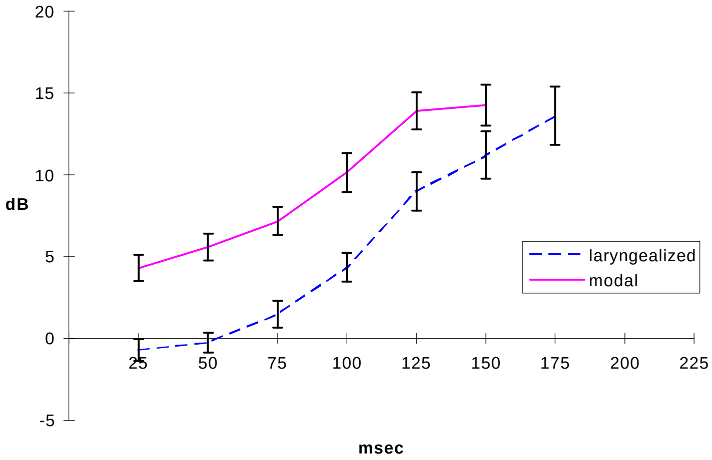
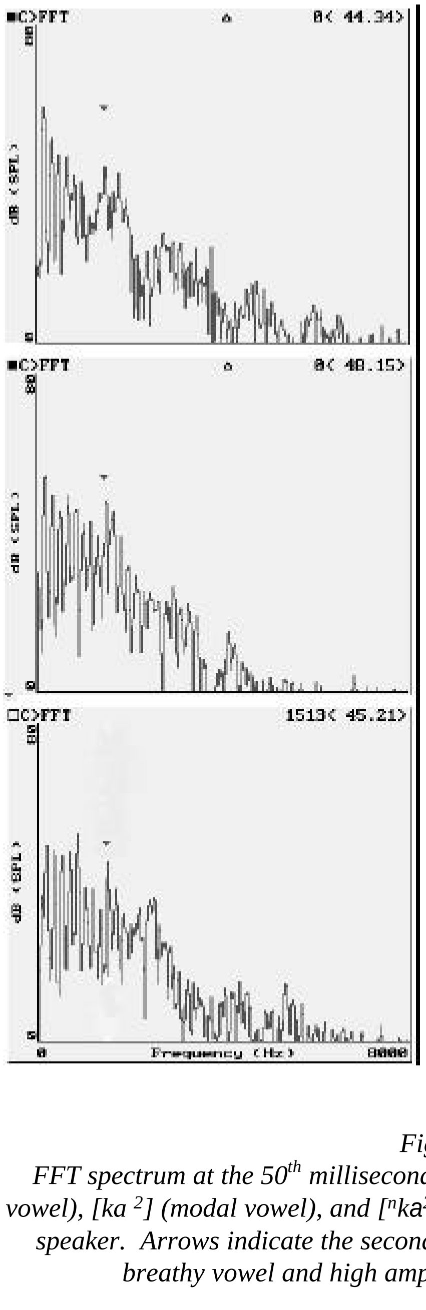


Figure 12.
MAZATEC, H1-H2, 12 speakers. Laryngealized vs. modal vowels.



[kiNa: ^{23}]

H1	60 dB
F2	44 dB

[ka ^{2}]

H1	55 dB
F2	48 dB

[^{n}ka^{2}]

H1	50 dB
F2	45 dB

Figure 13.
FFT spectrum at the 50th millisecond of the Mazatec words [kiNa: ^{23}] (breathy vowel), [ka ^{2}] (modal vowel), and [^{n}ka^{2}] (laryngealized vowel), spoken by the same speaker. Arrows indicate the second formant, which has low amplitude in the breathy vowel and high amplitude in the laryngealized vowel.

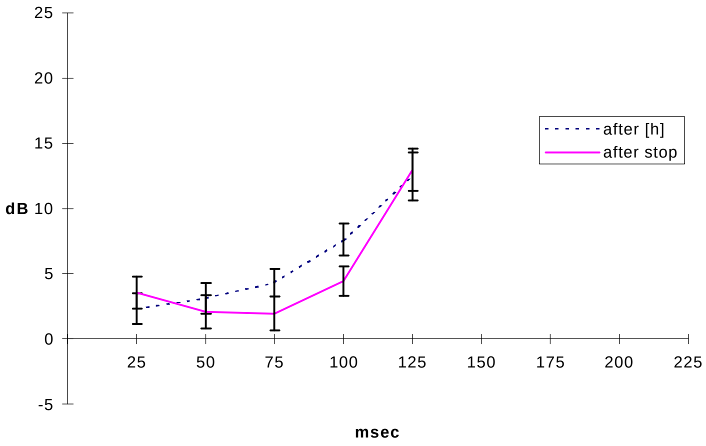


Figure 14.
TAGALOG, H1-F2, 12 speakers. Vowels after [h] vs. vowels after stop.

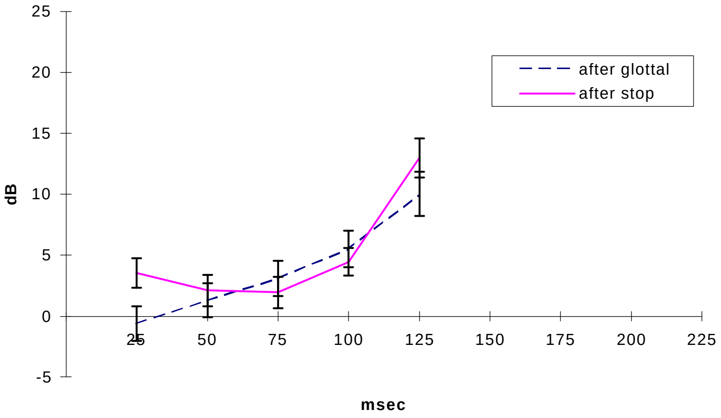


Figure 15.
TAGALOG, H1-F2, 12 speakers. Vowels after [/] vs. vowels after stop.

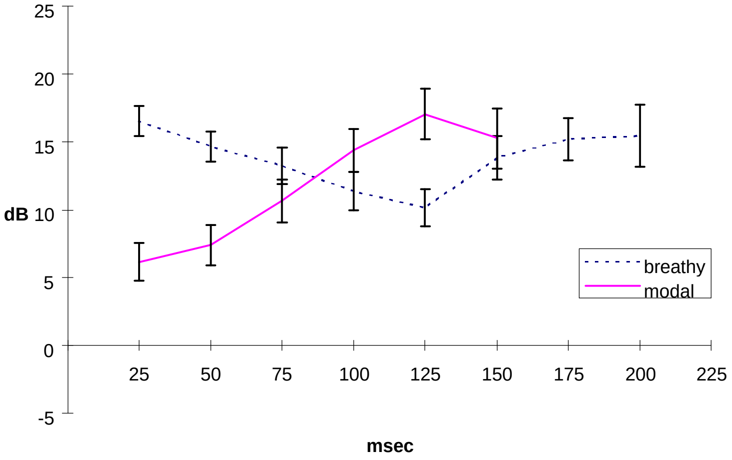


Figure 16.
MAZATEC, H1-F2, 12 speakers. Breathy vs. modal vowels.

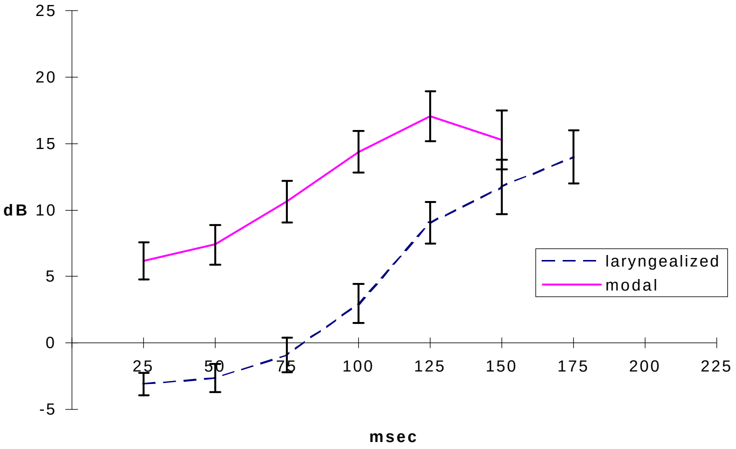


Figure 17.
MAZATEC, H1-F2, 12 speakers. Laryngealized vs. modal vowels.

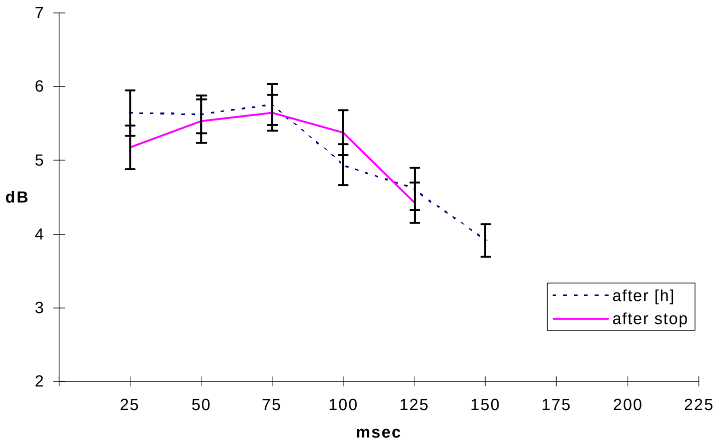


Figure 18.
TAGALOG, cepstral peak, 12 speakers. Vowels after [h] vs. vowels after stop.

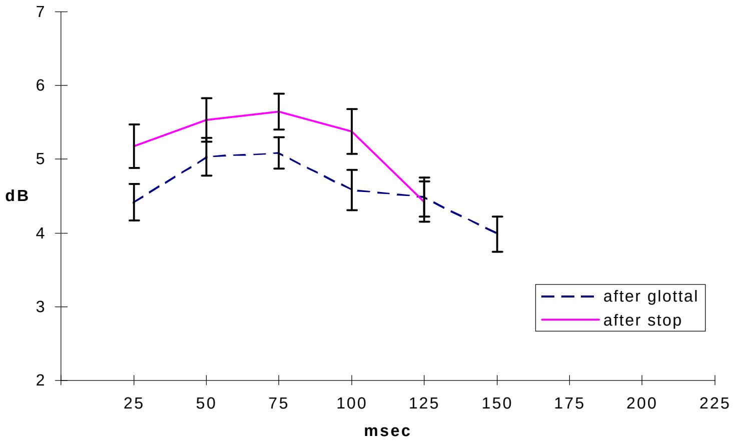


Figure 19.
TAGALOG, cepstral peak, 12 speakers. Vowels after [/] vs. vowels after stop.

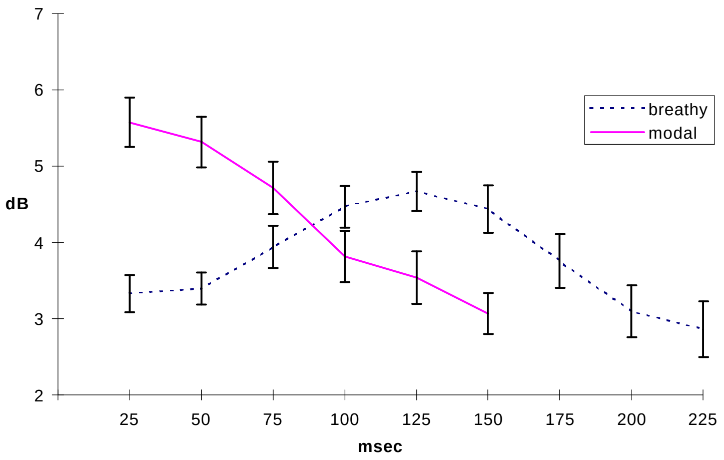


Figure 20.
MAZATEC, cepstral peak, 12 speakers. Breathy vs. modal vowels.

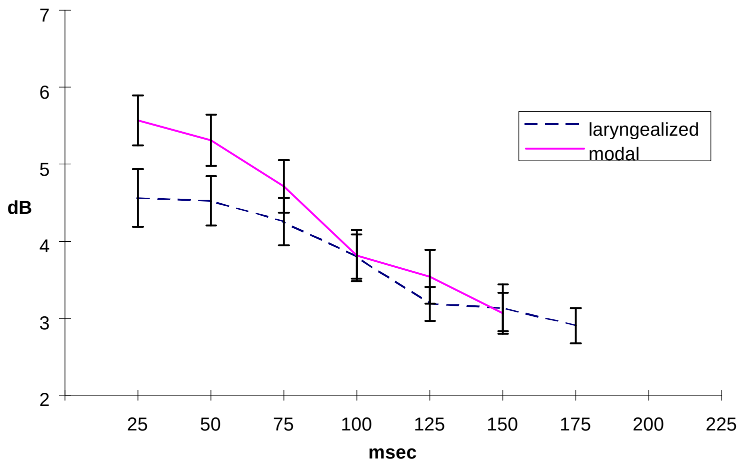


Figure 21.
MAZATEC, cepstral peak, 12 speakers. Laryngealized vs. modal vowels.

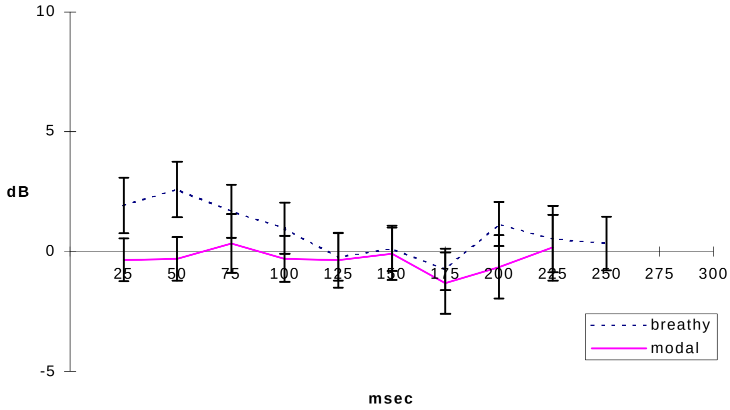


Figure 22.
CHONG, H1-H2, 8 speakers. Breathy vs. modal vowels.

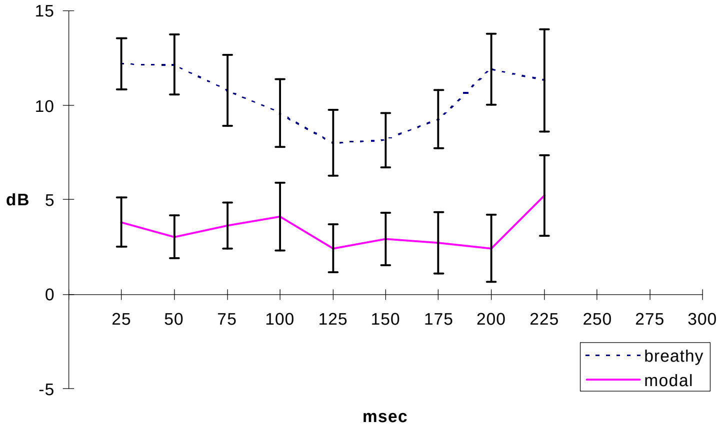


Figure 23.
CHONG, H1-F2, 8 speakers. Breathy vs. modal vowels.

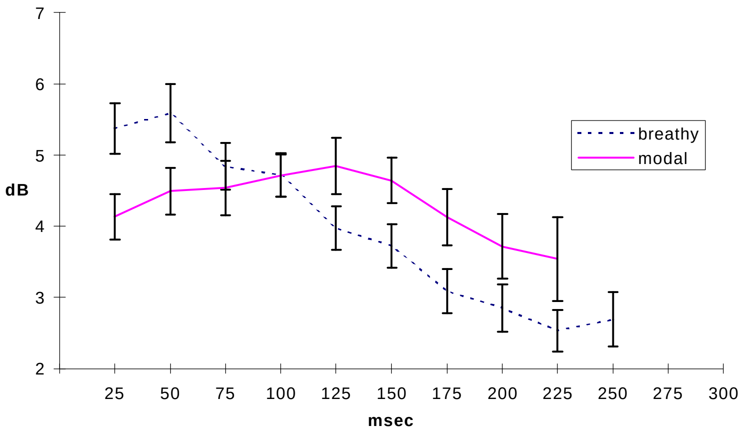


Figure 24.
CHONG, cepstral peak, 8 speakers. Breathy vs. modal vowels.

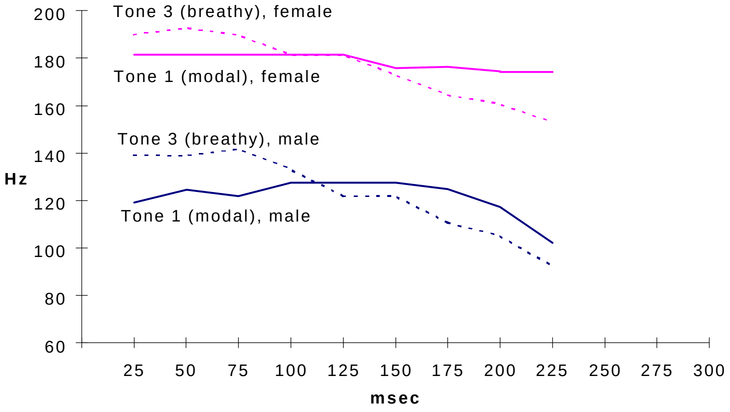


Figure 25.
CHONG, average pitch of F0, by tone and gender.
Tone 1 is modal and tone 3 is breathy.

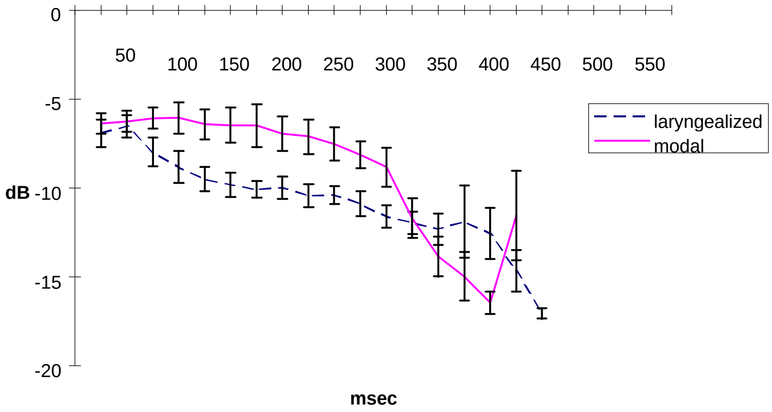


Figure 26.
MPI, H1-H2, 1 speaker. Laryngealized vs. modal vowels, high falling tone.
(9 samples in each group).

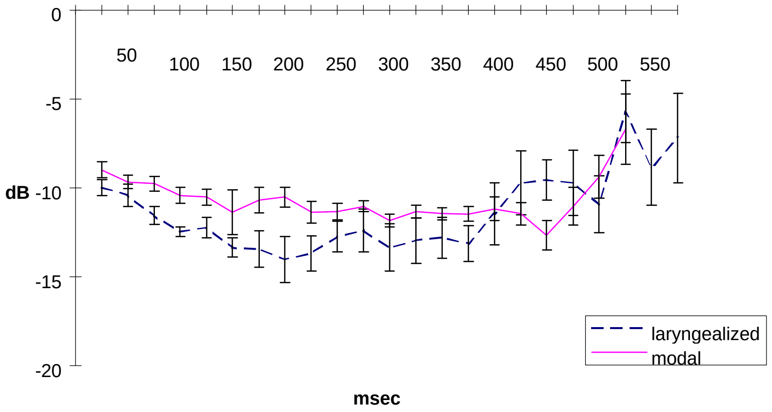


Figure 27.
MPI, H1-H2, 1 speaker. Laryngealized vs. modal vowels, mid rising tone.
(8 samples in each group).

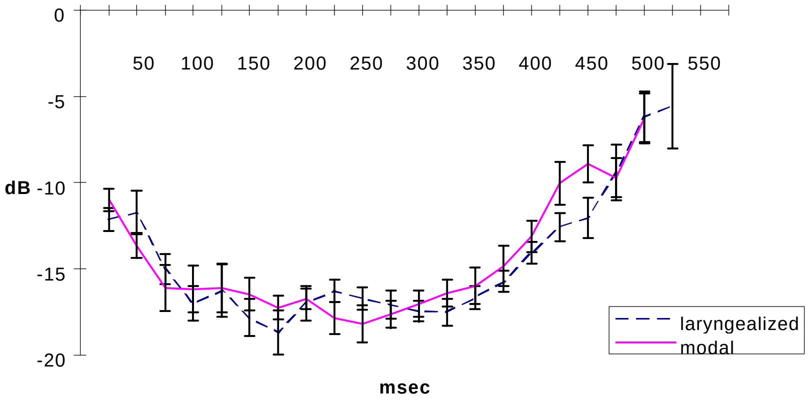


Figure 28.
MPI, H1-H2, 1 speaker. Laryngealized vs. modal vowels, low rising tone.
(9 samples in each group).

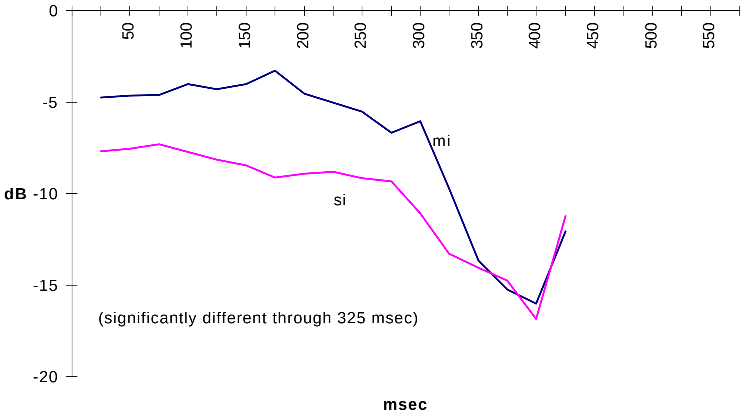


Figure 29.
MPI, H1-H2, 1 speaker. Modal vowels, high falling tone,
(mi, 4 iterations; si, 5 iterations).

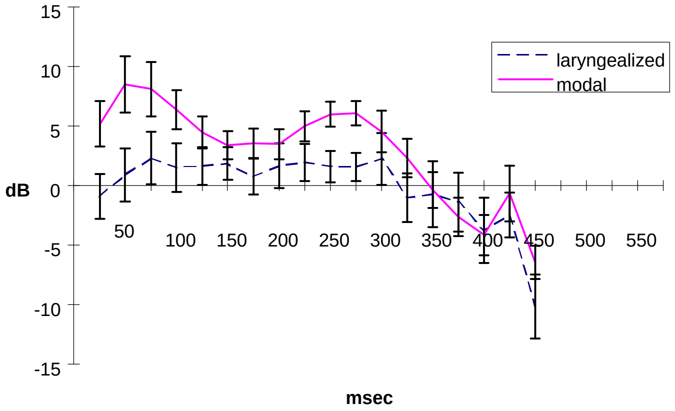


Figure 30.
MPI, H1-F2, 1 speaker. Laryngealized vs. modal vowels, high falling tone.
(9 samples in each group).

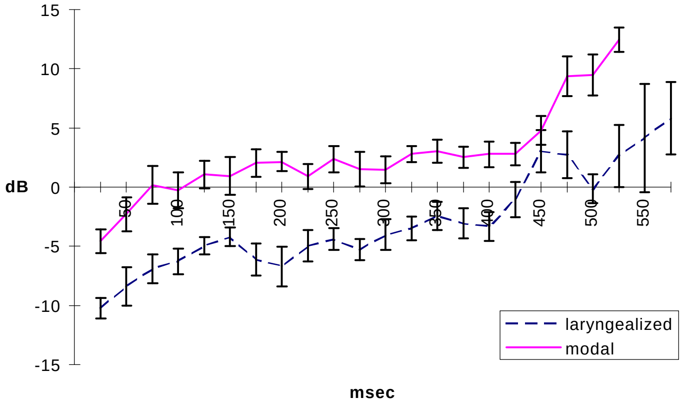


Figure 31.
MPI, H1-F2, 1 speaker. Laryngealized vs. modal vowels, mid rising tone.
(8 samples in each group).

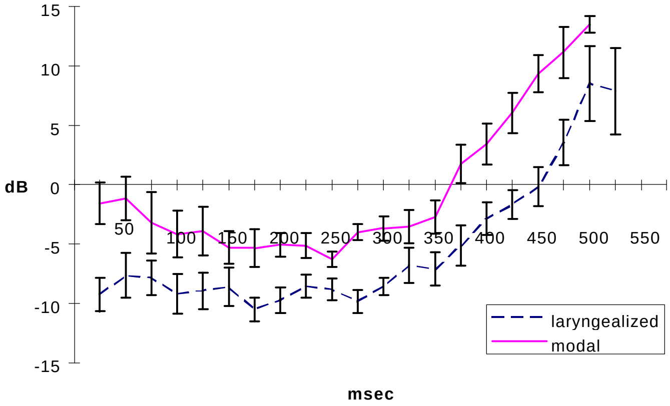


Figure 32.
MPI, H1-F2, 1 speaker. Laryngealized vs. modal vowels, low rising tone.
(9 samples in each group).

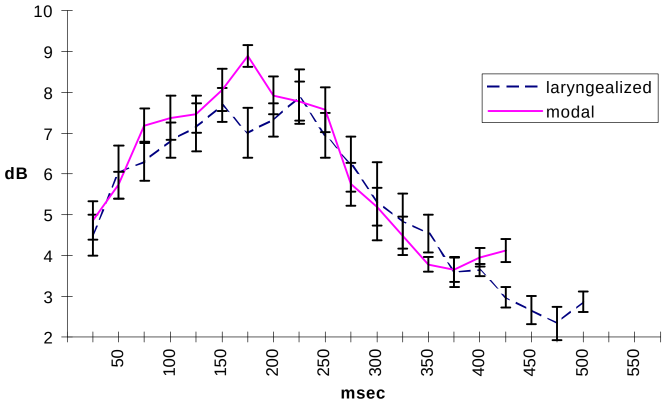


Figure 33.
MPI, Cepstral peak, 1 speaker. Laryngealized vs. modal vowels, high falling tone.
(9 samples in each group).

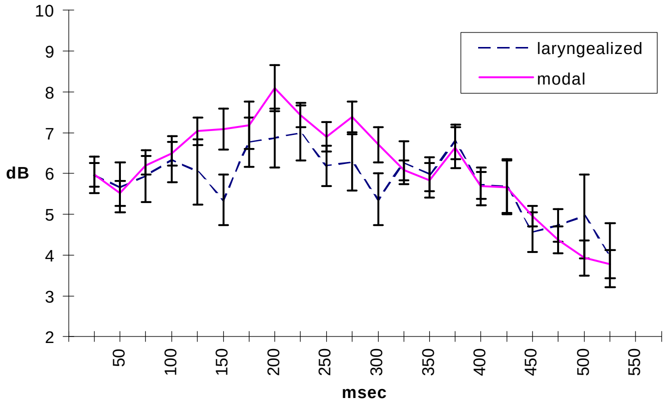


Figure 34.
MPI, Cepstral peak, 1 speaker. Laryngealized vs. modal vowels, mid rising tone.
(9 samples in each group).

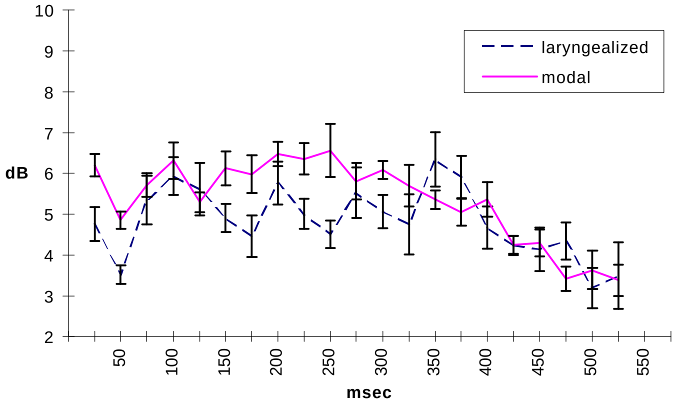


Figure 35.
MPI, Cepstral peak, 1 speaker. Laryngealized vs. modal vowels, low rising tone.
(9 samples in each group).

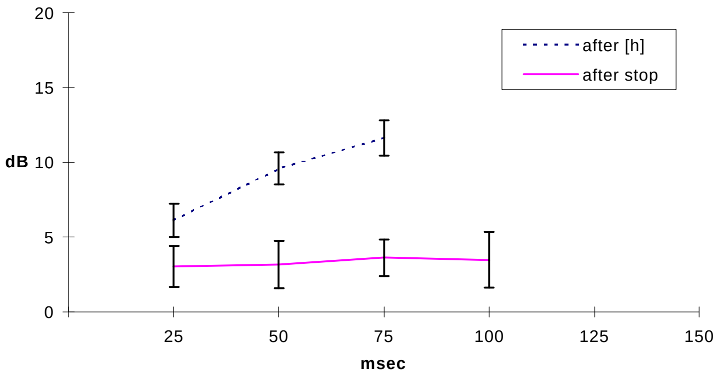


Figure 36.
NAVAJO, H1-H2, 11 speakers. Vowels after [h] vs. vowels after stop.

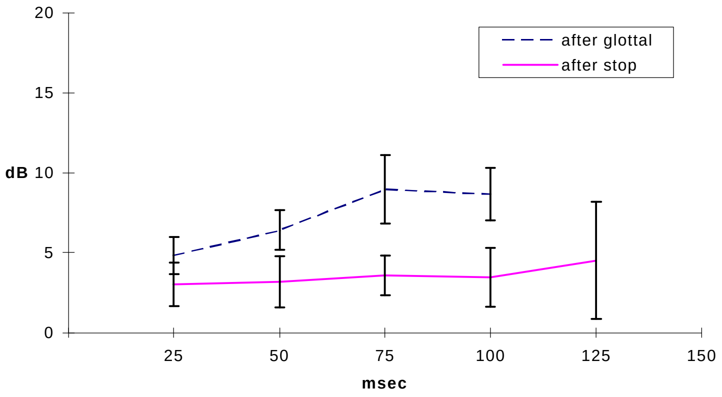


Figure 37.
NAVAJO, H1-H2, 11 speakers. Vowels after [/] vs. vowels after stop.

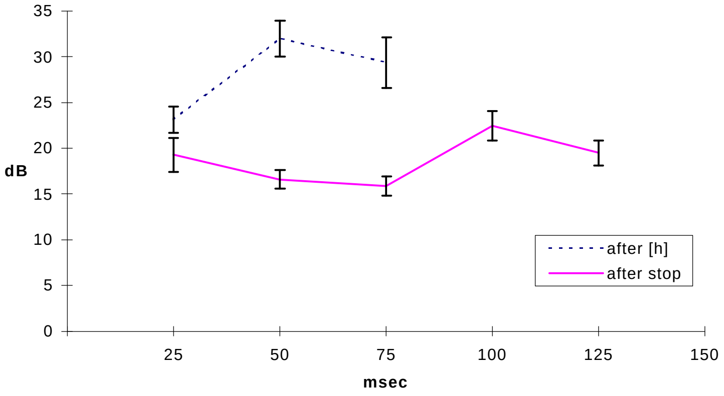


Figure 38.
NAVAJO, H1-F2, 11 speakers. Vowels after [h] vs. vowels after stop.

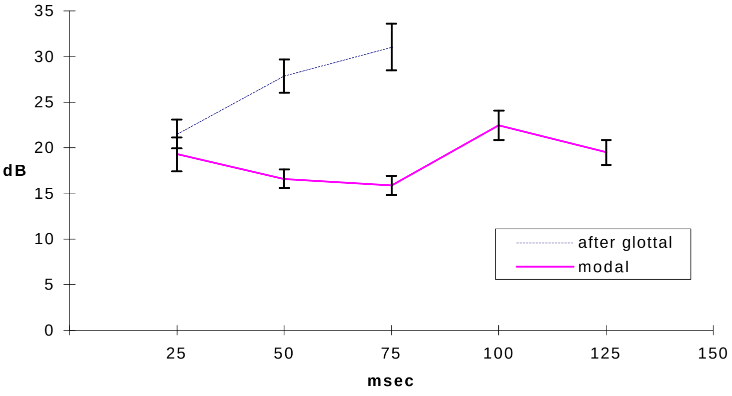


Figure 39.
NAVAJO, H1-F2, 11 speakers. Vowels after [/] vs. vowels after stop.

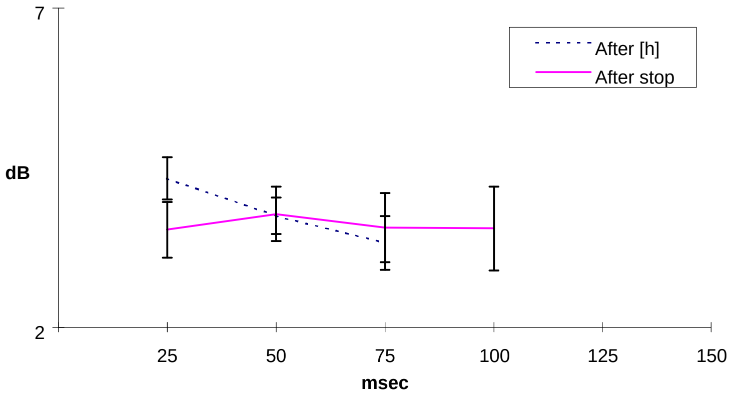


Figure 40.
NAVAJO, Cepstral peak, 11 speakers. Vowels after [h] vs. vowels after stop.

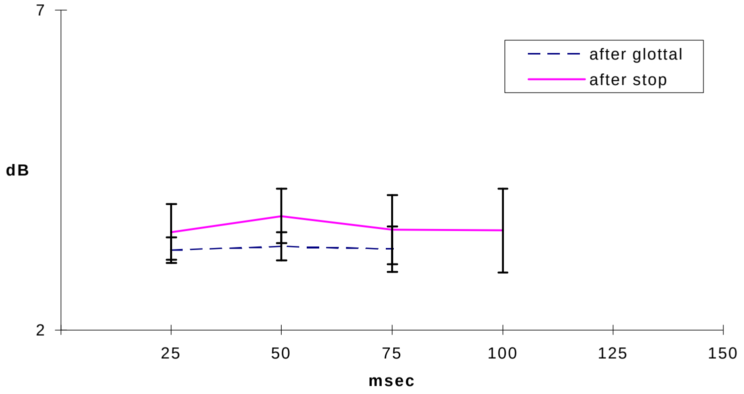


Figure 41.
NAVAJO, Cepstral peak, 11 speakers. Vowels after [/] vs. vowels after stop.

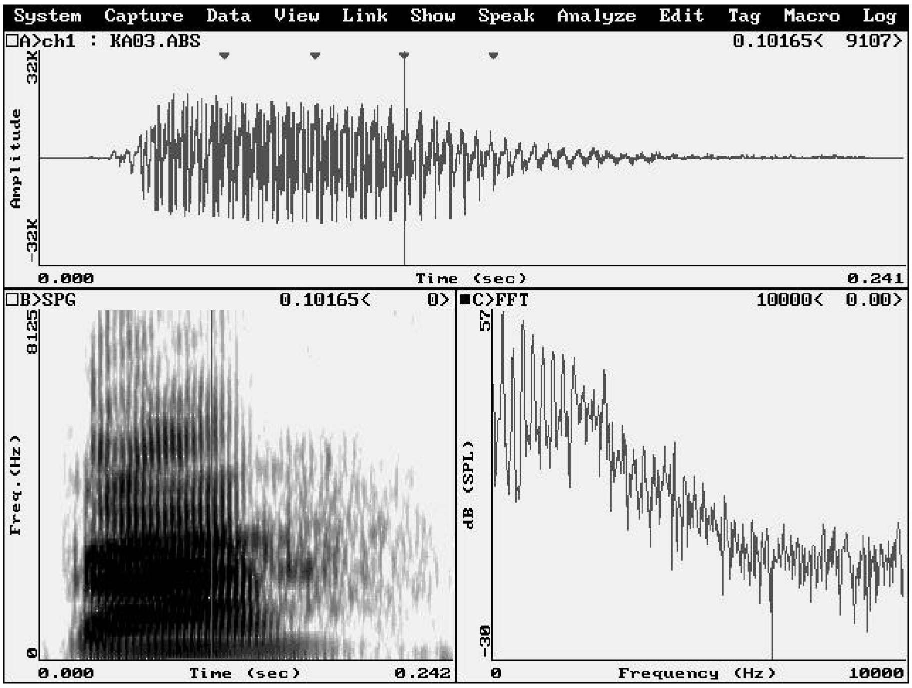


Figure 42.
A Mazatec modal vowel in the word [ka ^{2}] 'bald', speaker 3, female.

Window A: Waveform

Window B: Spectrogram of the waveform.

Window C: Spectrum of the waveform, taken over a 25 ms interval centered at the 75th millisecond, the location marked by the cursor in window A.

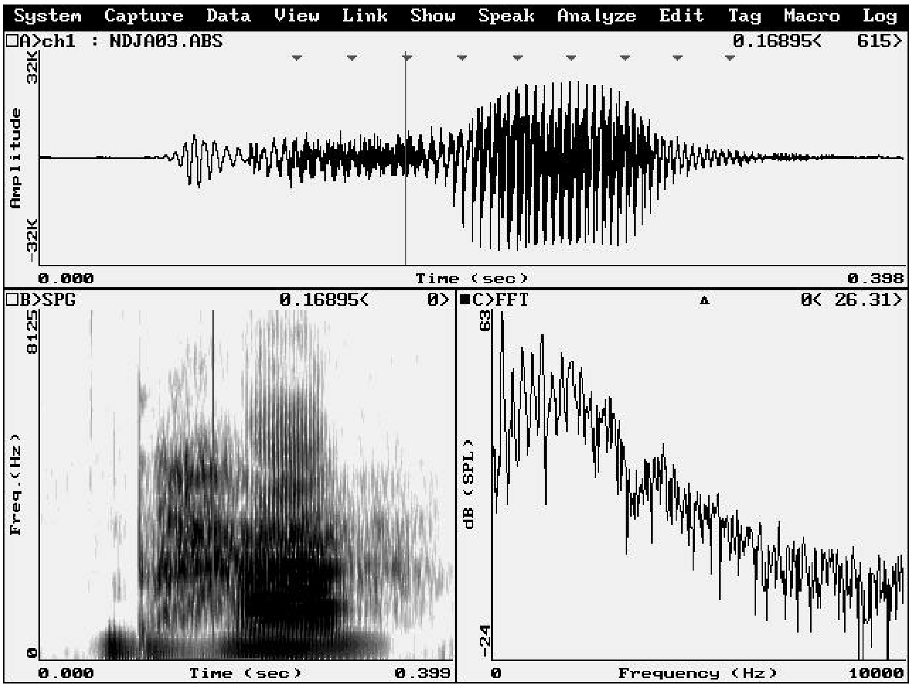


Figure 43.
A Mazatec breathy vowel in the word [^{n}da: ^{23}] 'cornflour drink', speaker 3, female.

Window A: Waveform

Window B: Spectrogram of the waveform.

Window C: Spectrum of the waveform, taken over a 25 ms interval centered at the 75th millisecond, the location marked by the cursor in window A. The differences H1-H2 and F1-F2 are greater than in the modal spectrum of figure 42.

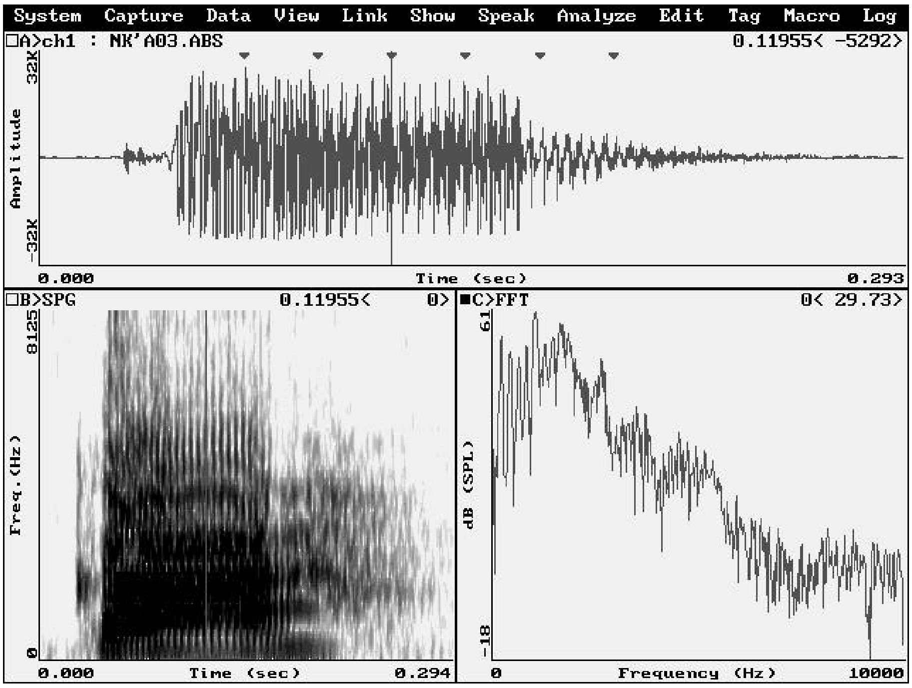


Figure 44.
A Mazatec laryngealized vowel in the word [^{n}ka^{2}] ‘high’, speaker 3, female.

Window A: Waveform

Window B: Spectrogram of the waveform. There is no visible creak.

Window C: Spectrum of the waveform, taken over a 25 ms interval centered at the 75th millisecond, the location marked by the cursor in window A. The differences H1-H2 and F1-F2 are less than in the modal spectrum of figure 42.

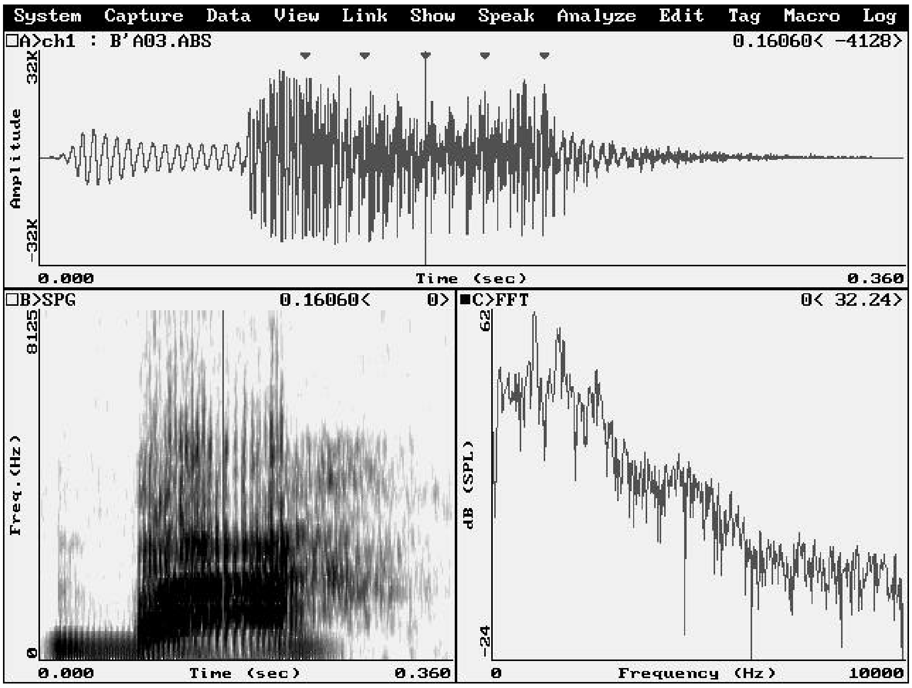


Figure 45.
A Mazatec laryngealized vowel in the word [Ba^{2}] ‘carries’, speaker 3, female.

Window A: Waveform

Window B: Spectrogram of the waveform. There is visible creak.

Window C: Spectrum of the waveform, taken over a 25 ms interval centered at the 75th millisecond, the location marked by the cursor in window A. The differences H1-H2 and F1-F2 are less than in the modal spectrum of figure 42, and similar to those in the laryngealized spectrum of figure 44.

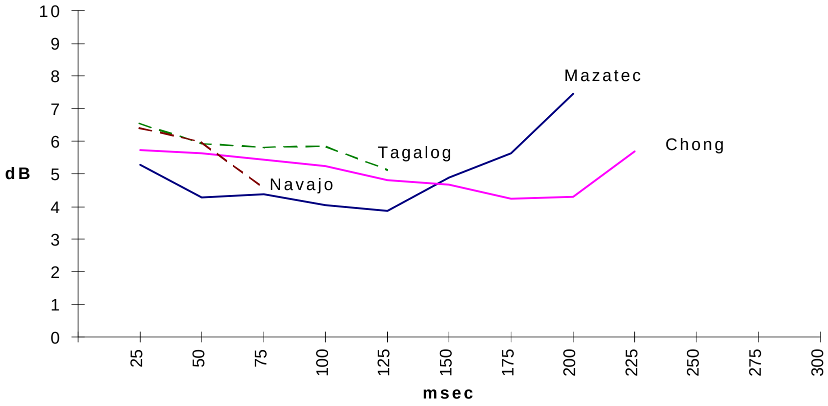


Figure 46.
Standard deviation, H1-H2. Breathy vowels.

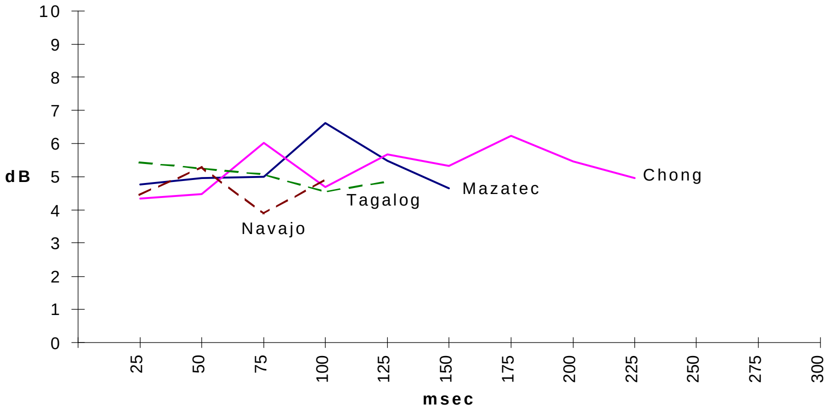


Figure 47.
Standard deviation, H1-H2. Modal vowels.

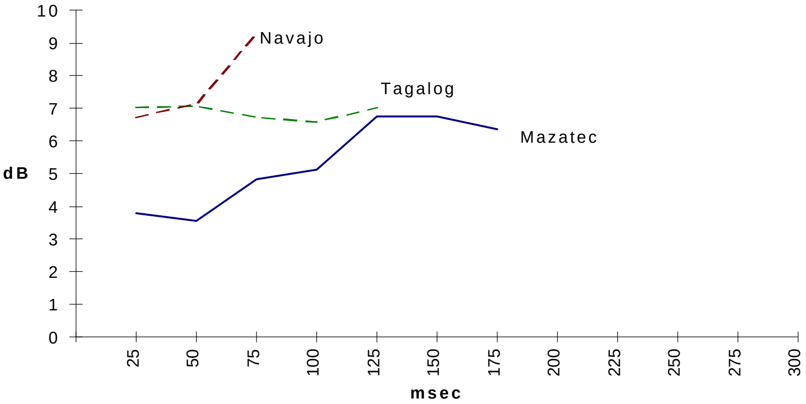


Figure 48.
Standard deviation, H1-H2. Laryngealized vowels.

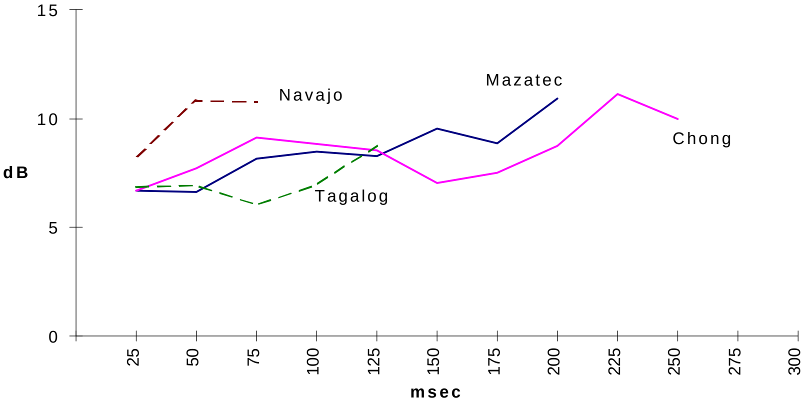


Figure 49.
Standard deviation, H1-F2. Breathy vowels.

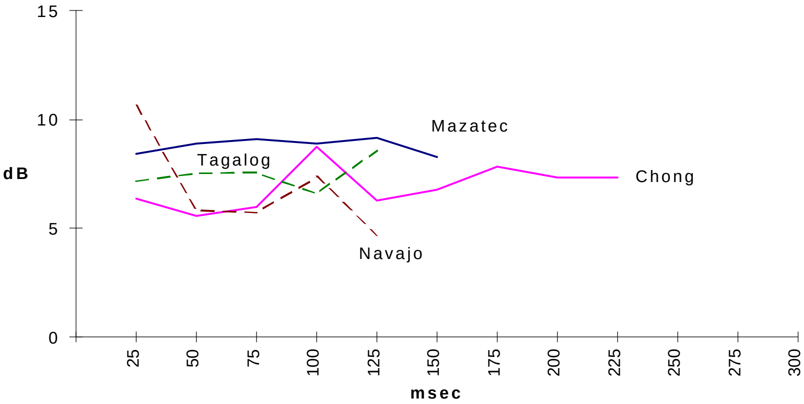


Figure 50.
Standard deviation, H1-F2. Modal vowels.

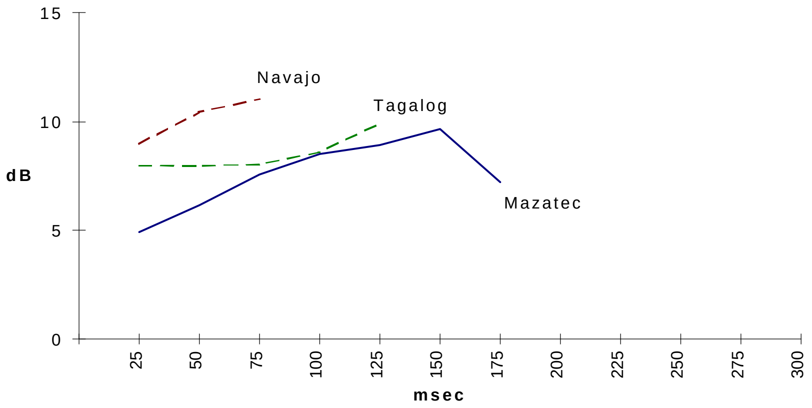


Figure 51.
Standard deviation, H1-F2. Laryngealized vowels.

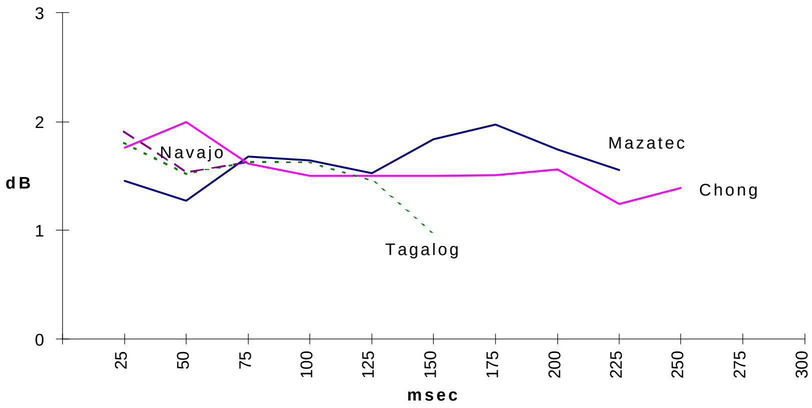


Figure 52.
Standard deviation, cepstral peak. Breathy vowels.

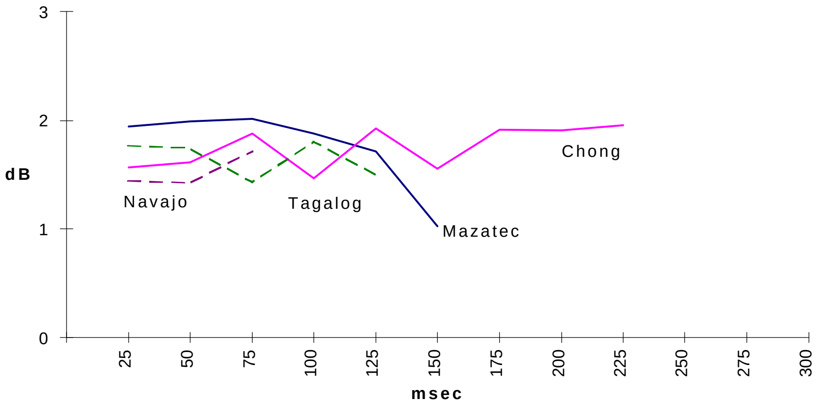


Figure 53.
Standard deviation, cepstral peak. Modal vowels.

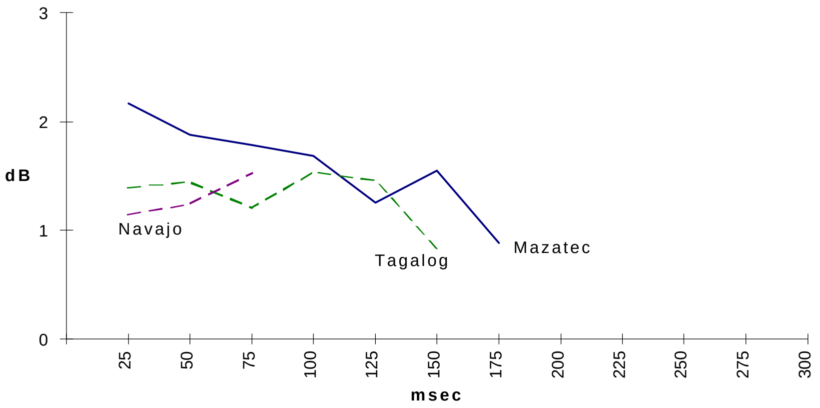


Figure 54.
Standard deviation, cepstral peak. Laryngealized vowels.

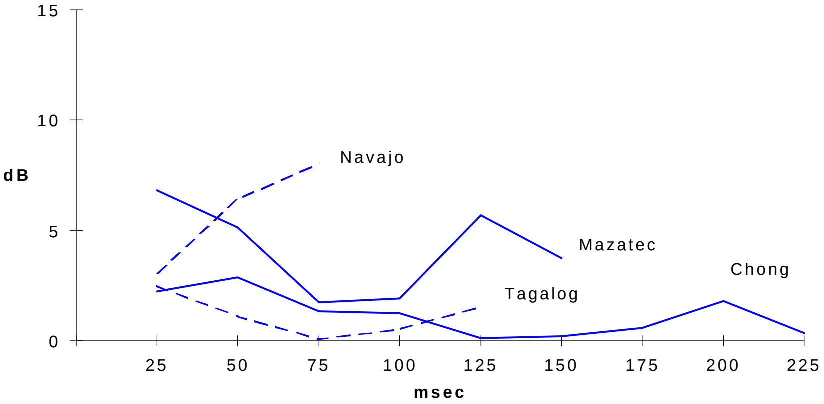


Figure 55.
Difference between breathy and modal vowels.
H1-H2 measurement, absolute values.

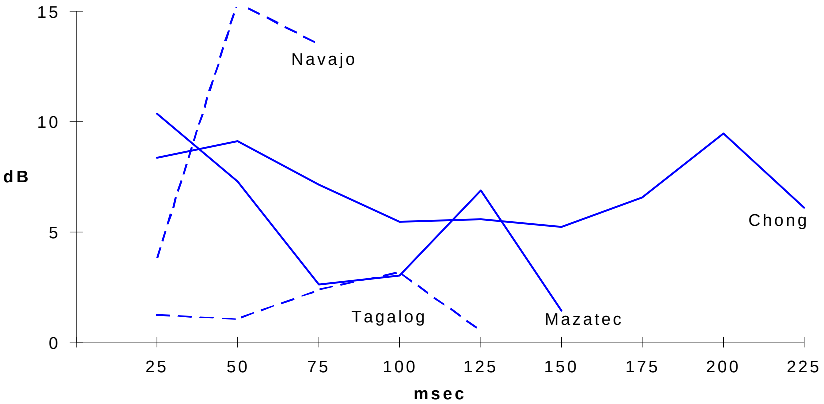


Figure 56.
Difference between breathy and modal vowels.
H1-F2 measurement, absolute values.

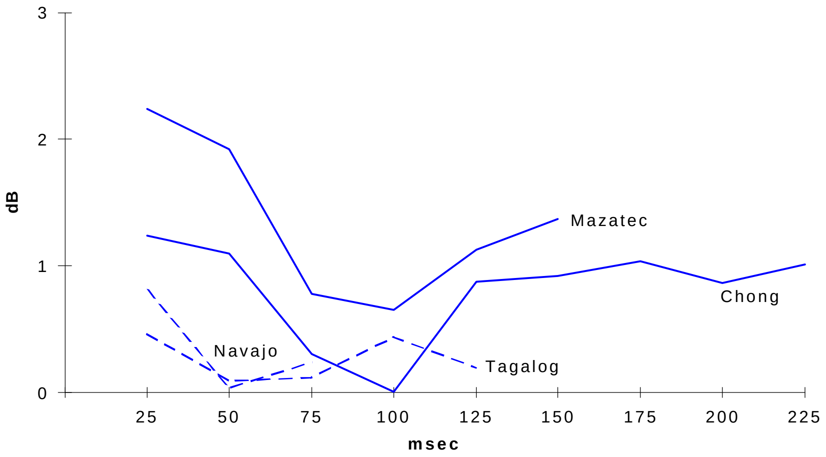


Figure 57.
Difference between breathy and modal vowels.
Cepstral peak measurement, absolute values.

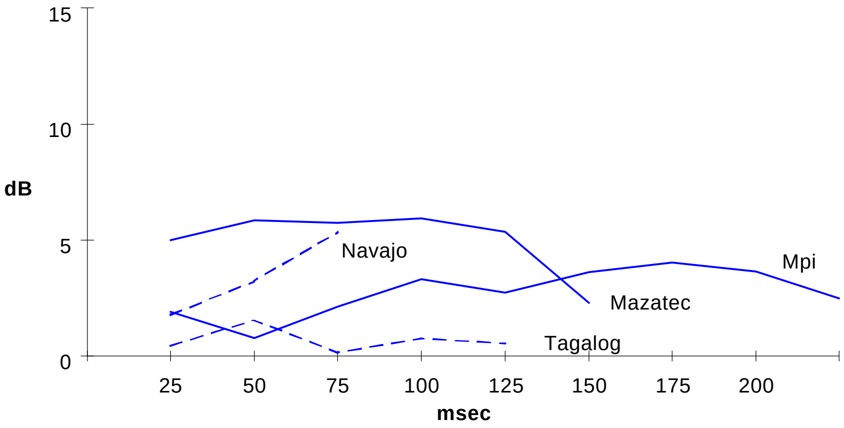


Figure 58.
Difference between laryngealized and modal vowels.
H1-H2 measurement, absolute values.

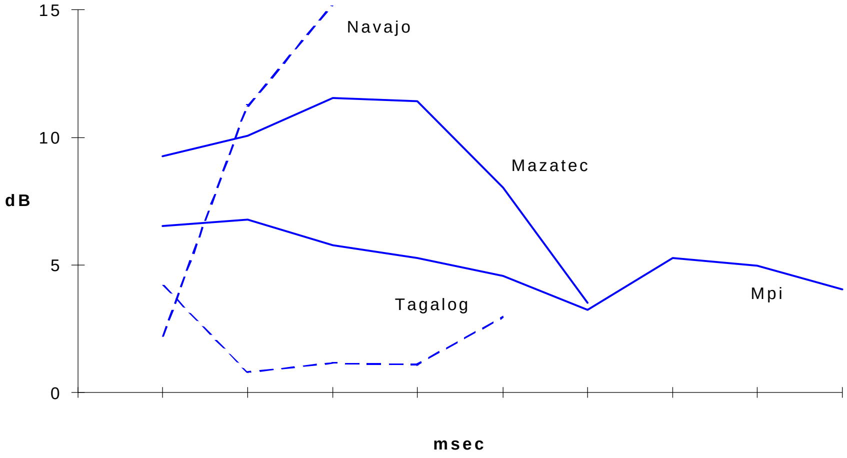


Figure 59.
Difference between laryngealized and modal vowels.
H1-F2 measurement, absolute values.

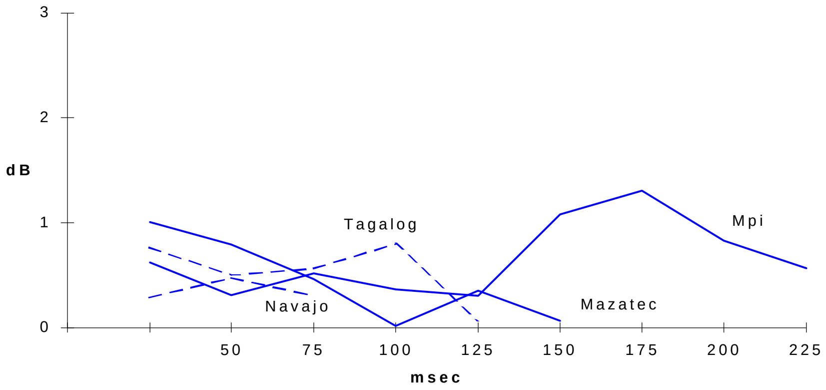


Figure 60.
Difference between laryngealized and modal vowels.
Cepstral peak measurement, absolute values.

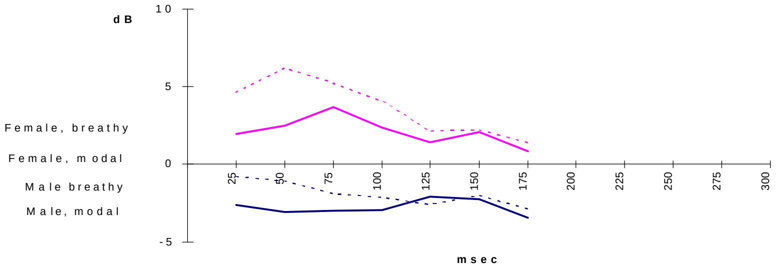


Figure 61.
CHONG, H1-H2, 8 speakers. Gender and type differences.

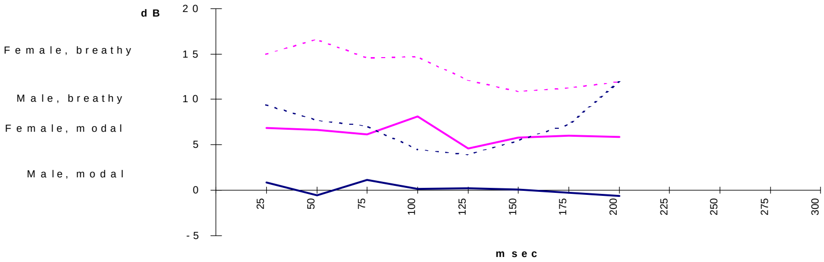


Figure 62.
CHONG, H1-F2, 8 speakers. Gender and type differences.

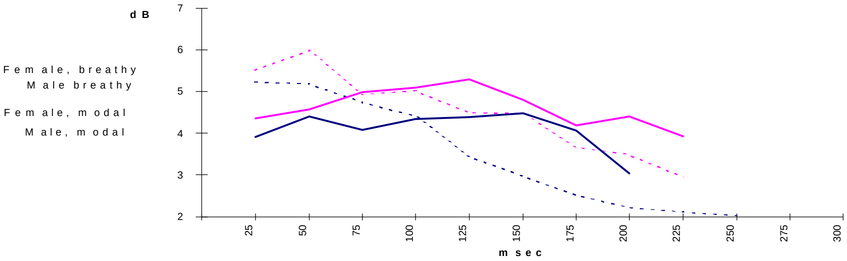


Figure 63.
CHONG, cepstral peak, 8 speakers. Gender and type differences.

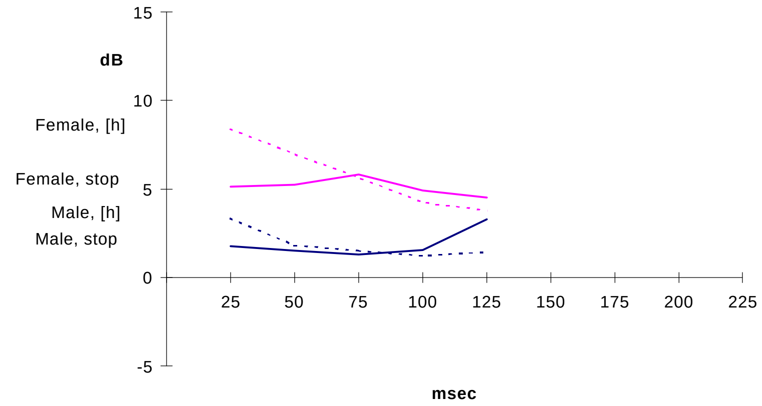


Figure 64.
TAGALOG, H1-H2, 12 speakers. Gender differences.
Vowels after [h] vs. vowels after stop.

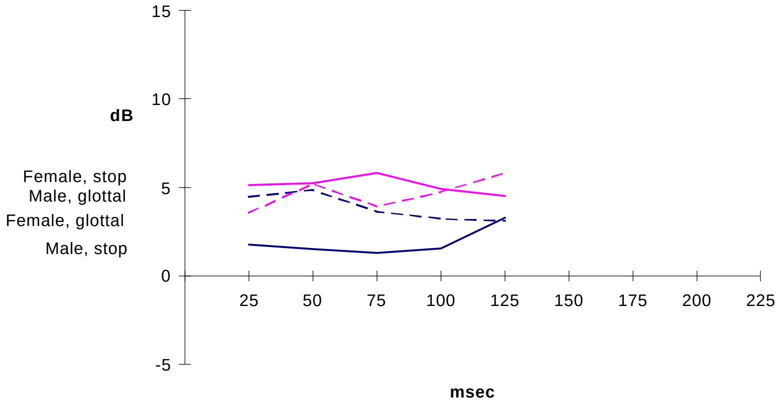


Figure 65.
TAGALOG, H1-H2, 12 speakers. Gender differences.
Vowels after glottal [/] vs. vowels after stop.

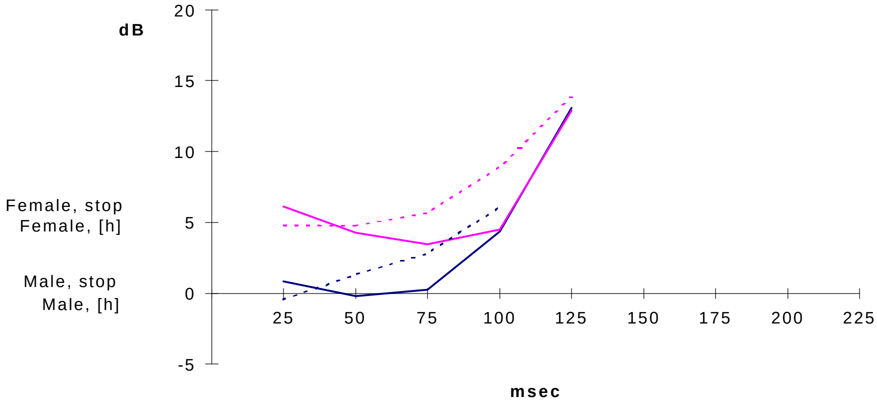


Figure 66.
TAGALOG, H1-F2, 12 speakers. Gender differences.
Vowels after [h] vs. vowels after stop.

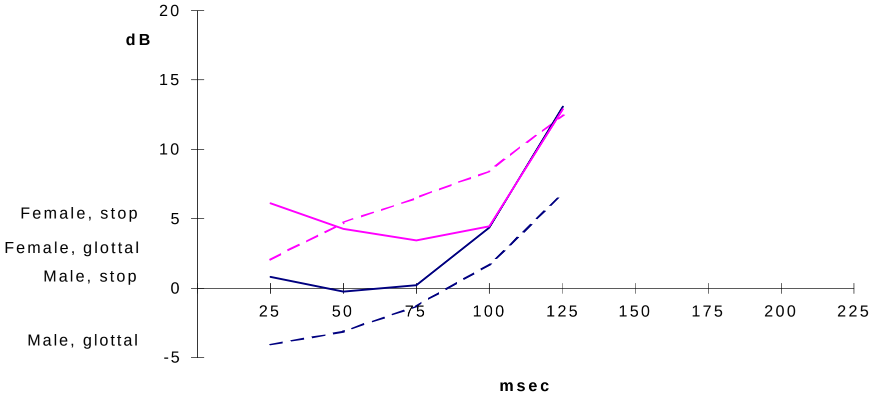


Figure 67.
TAGALOG, H1-F2, 12 speakers. Gender differences.
Vowels after glottal [/] vs. vowels after stop.

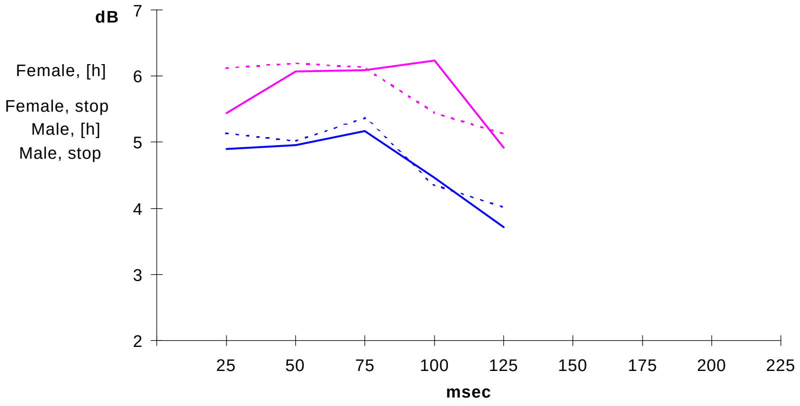


Figure 68.
TAGALOG, cepstral peak, 12 speakers. Gender differences.
Vowels after [h] vs. vowels after stop.

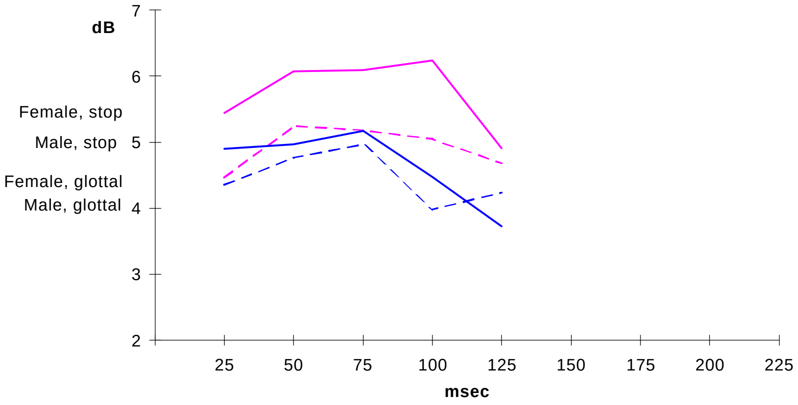


Figure 69.
TAGALOG, cepstral peak, 12 speakers. Gender differences.
After glottal [/] vs. vowels after stop.

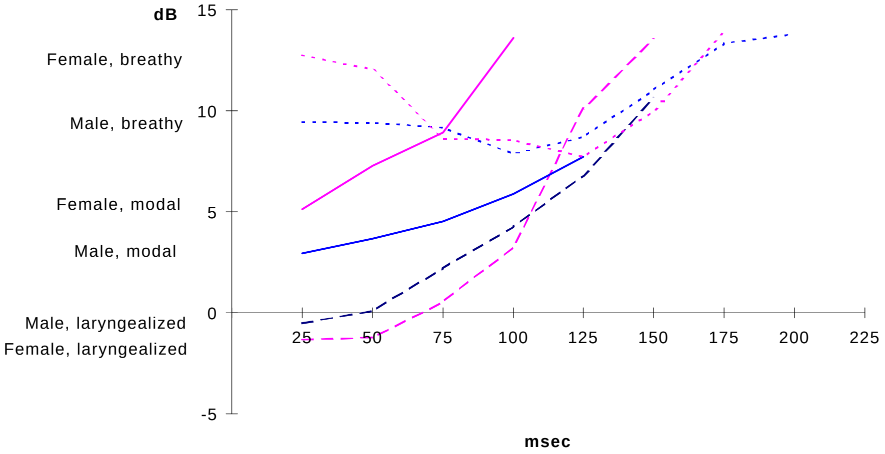


Figure 70.
MAZATEC, H1-H2, 12 speakers. Gender and type differences.

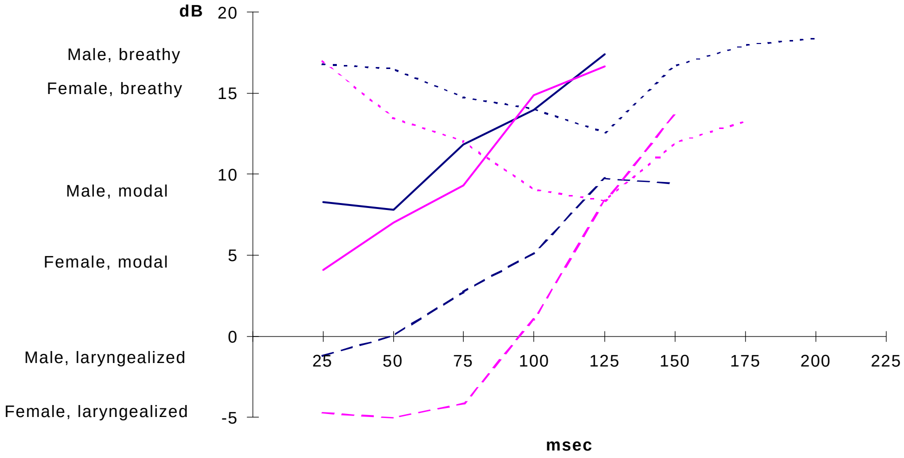


Figure 71.
MAZATEC, H1-F2, 12 speakers. Gender and type differences.

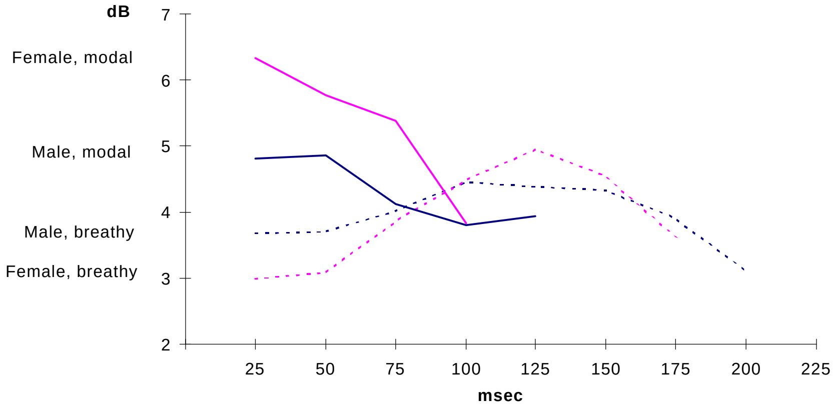


Figure 72.
MAZATEC, cepstral peak, 12 speakers. Gender differences.
Breathy vs. modal vowels.

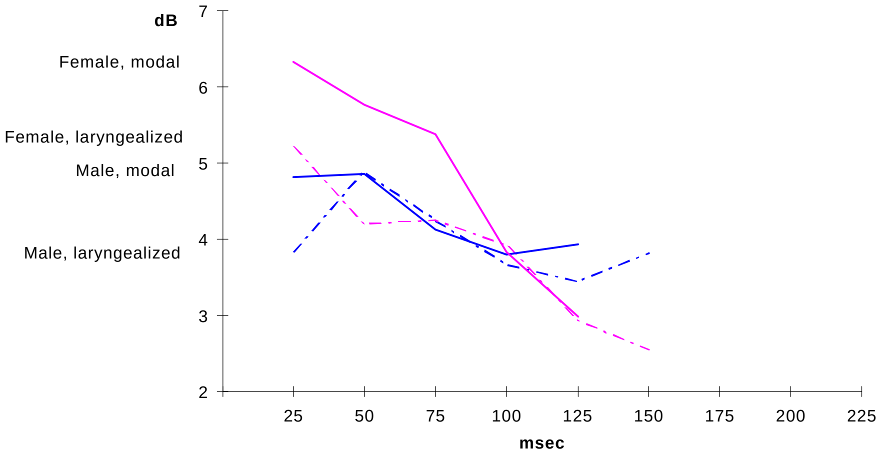


Figure 73.
MAZATEC, cepstral peak, 12 speakers. Gender differences.
Laryngealized vs. modal vowels.

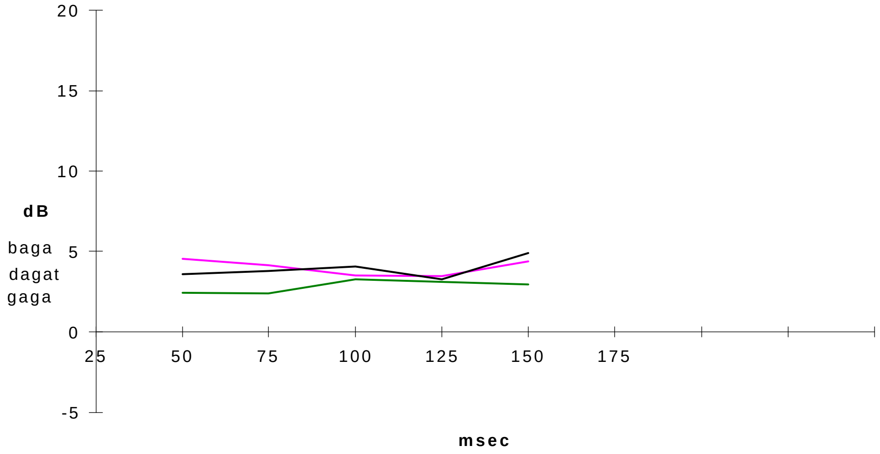


Figure 74.
TAGALOG, H1-H2, 12 speakers. Vowels after stop.

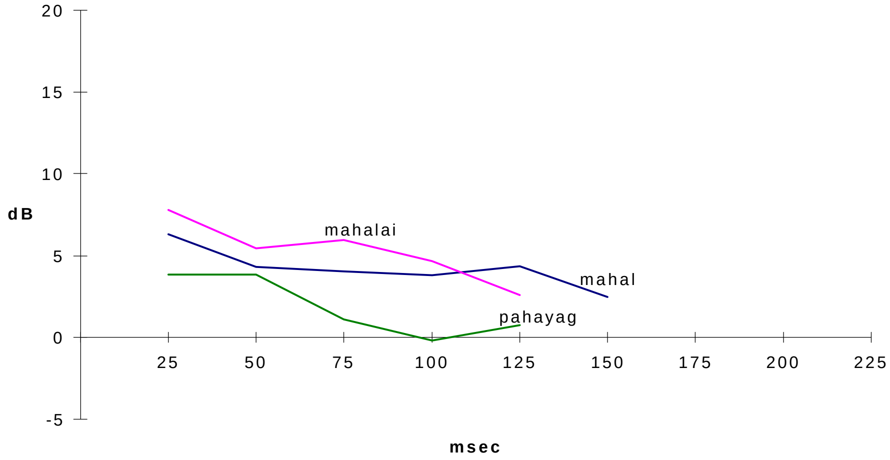


Figure 75.
TAGALOG, H1-H2, 12 speakers. Vowels after [h].

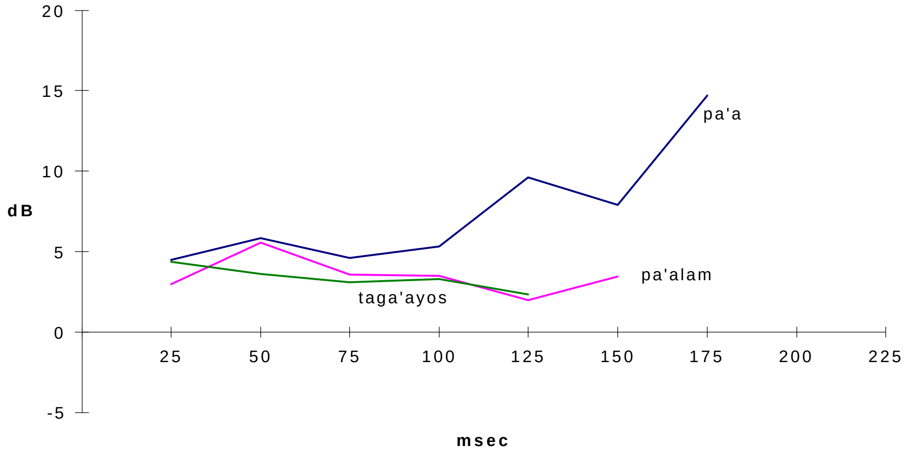


Figure 76.
TAGALOG, H1-H2, 12 speakers. Vowels after [/].

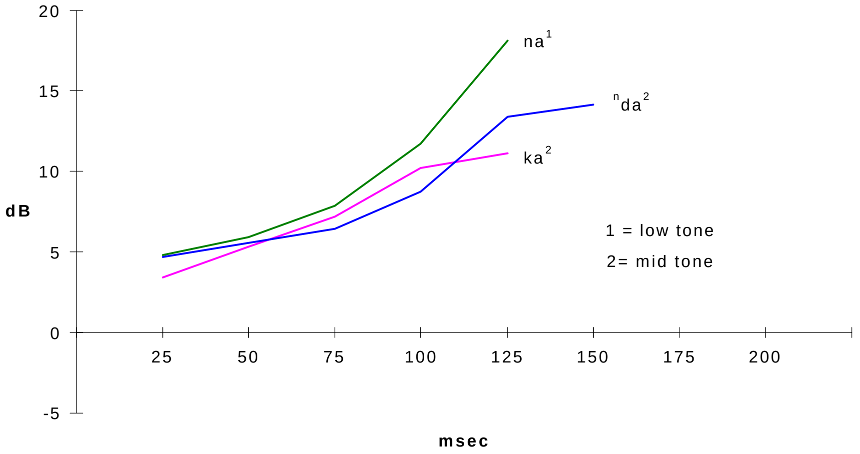


Figure 77.
MAZATEC, H1-H2, 12 speakers. Modal vowels.

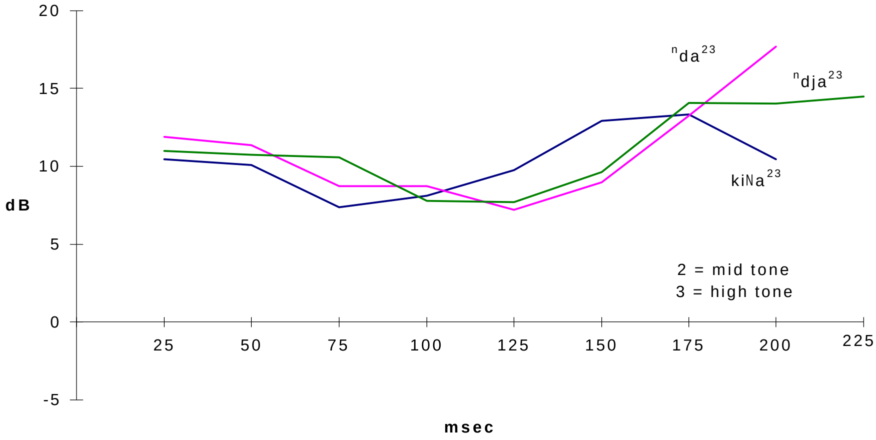


Figure 78.
MAZATEC, H1-H2, 12 speakers. Breathy vowels.

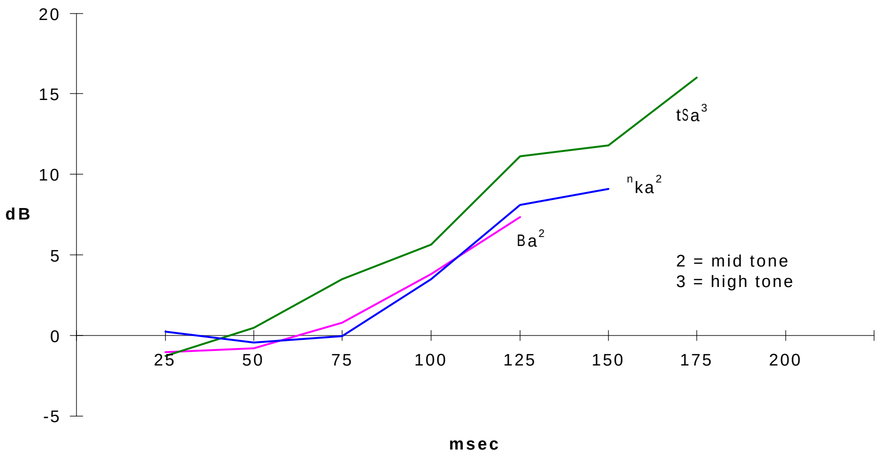


Figure 79.
MAZATEC, H1-H2, 12 speakers. Laryngealized vowels.

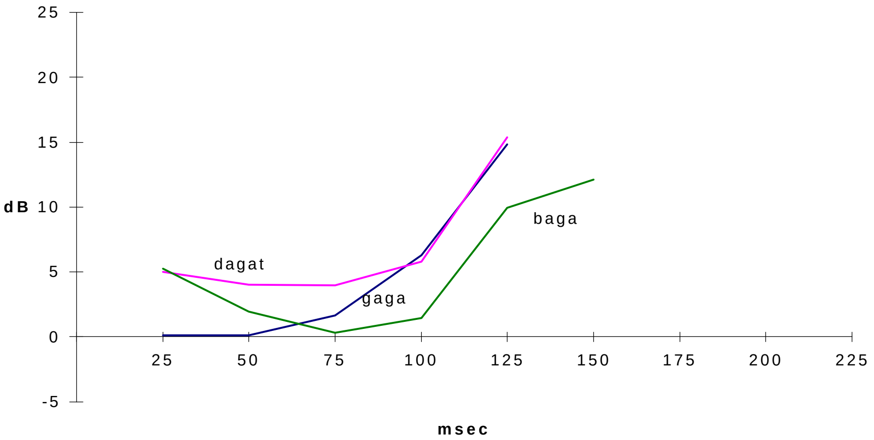


Figure 80.
TAGALOG, H1-F2, 12 speakers. Vowels after stop.

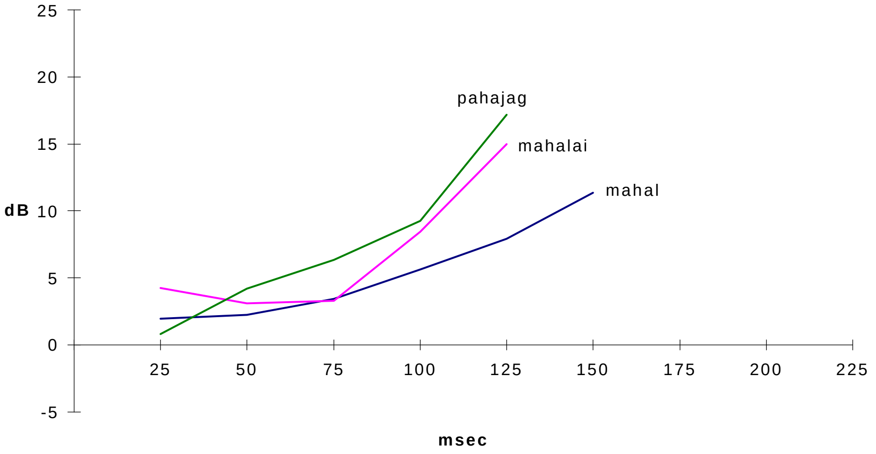


Figure 81.
TAGALOG, H1-F2, 12 speakers. Vowels after [h].

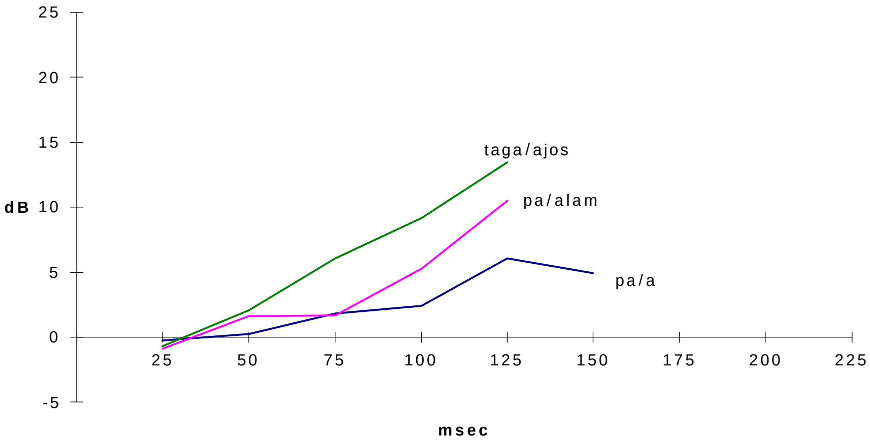


Figure 82.
TAGALOG, H1-F2, 12 speakers. Vowels after [/].

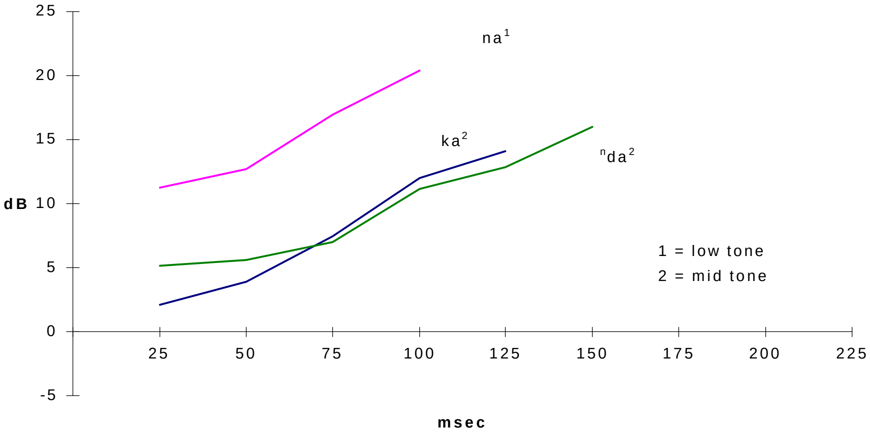


Figure 83.
MAZATEC, H1-F2, 12 speakers. Modal vowels.

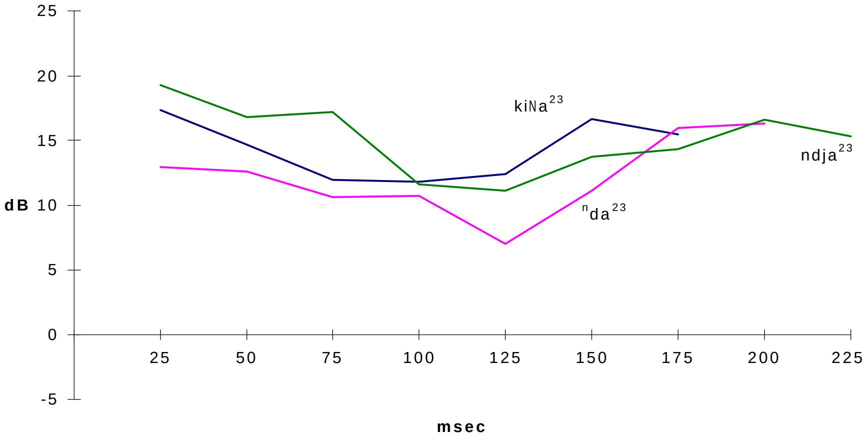


Figure 84.
MAZATEC, H1-F2, 12 speakers. Breathy vowels.

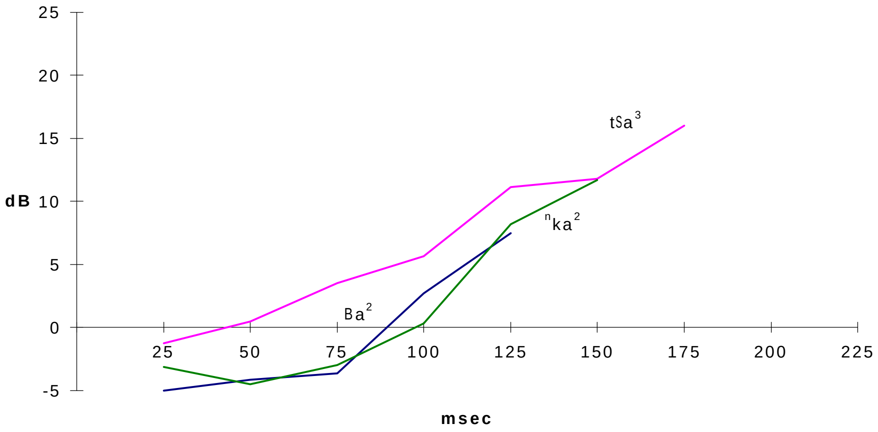


Figure 85.
MAZATEC, H1-F2, 12 speakers. Laryngealized vowels.

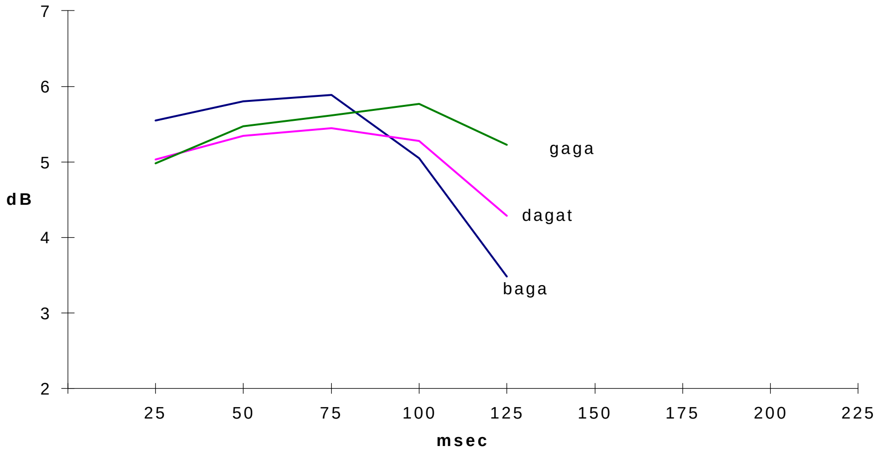


Figure 86.
TAGALOG, cepstral peak, 12 speakers. Vowels after stop.

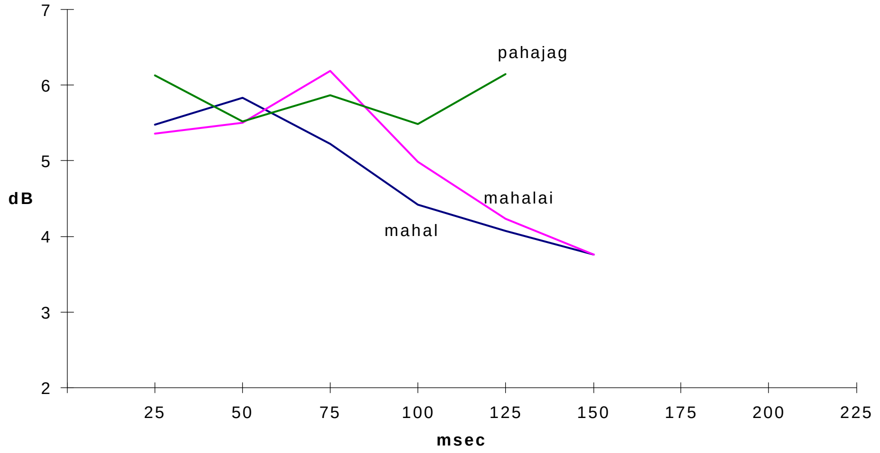


Figure 87.
TAGALOG, cepstral peak, 12 speakers. Vowels after [h].

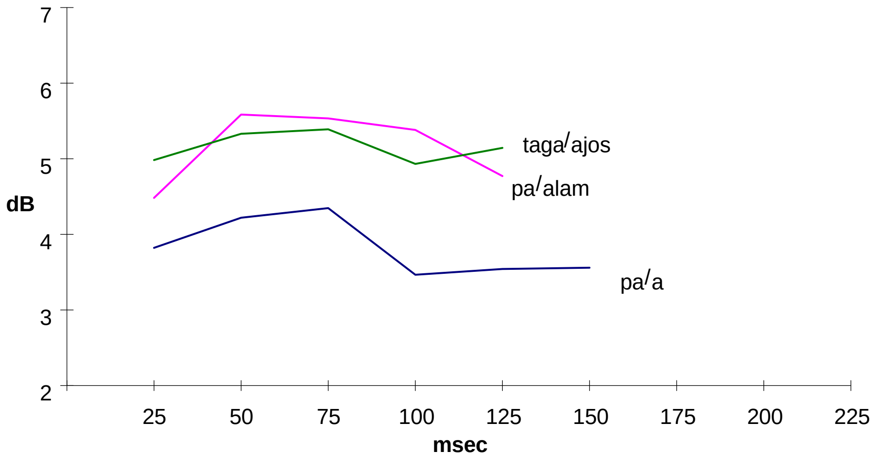


Figure 88.
TAGALOG, cepstral peak, 12 speakers. Vowels after [/].

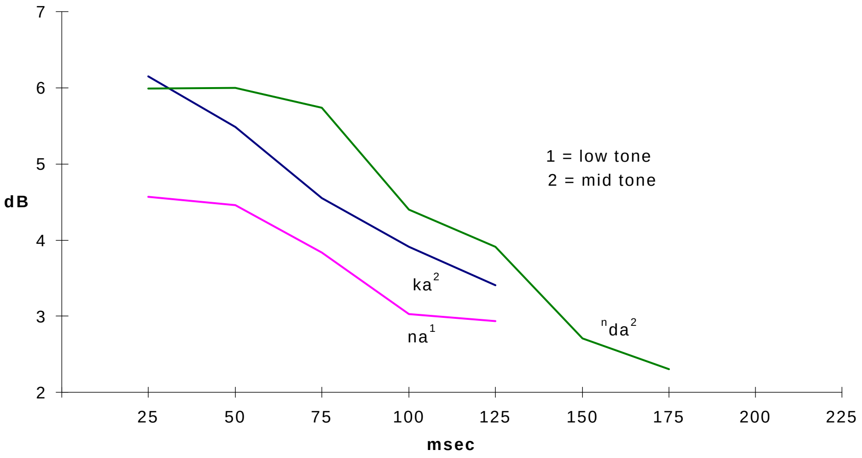


Figure 89.
MAZATEC, cepstral peak, 12 speakers. Modal vowels.

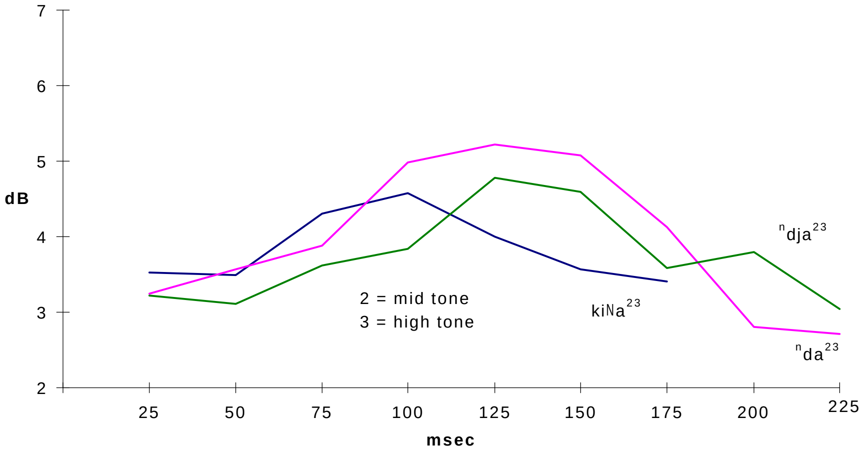


Figure 90.
MAZATEC, cepstral peak, 12 speakers. Breathy vowels.

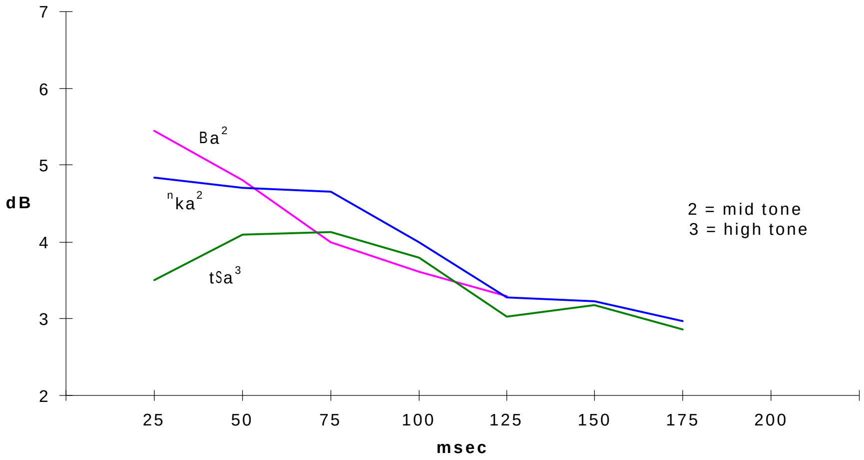


Figure 91.
MAZATEC, cepstral peak, 12 speakers. Laryngealized vowels.

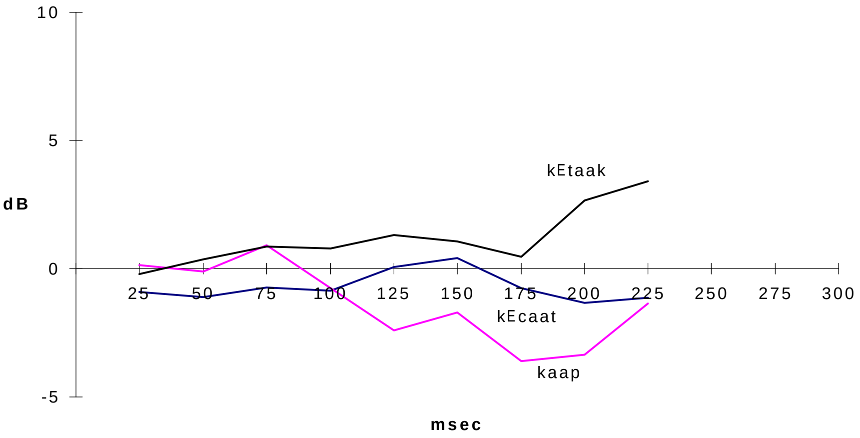


Figure 92.
CHONG, H1-H2, 8 speakers. Modal vowels.

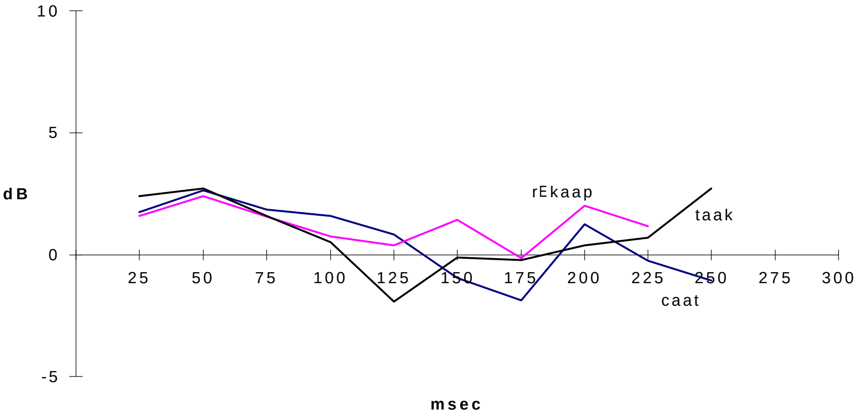


Figure 93.
CHONG, H1-H2, 8 speakers. Breathy vowels.

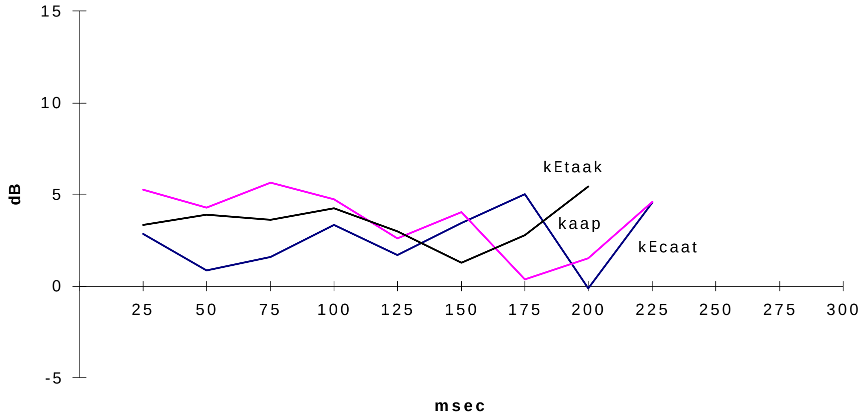


Figure 94.
CHONG, H1-F2, 8 speakers. Modal vowels.

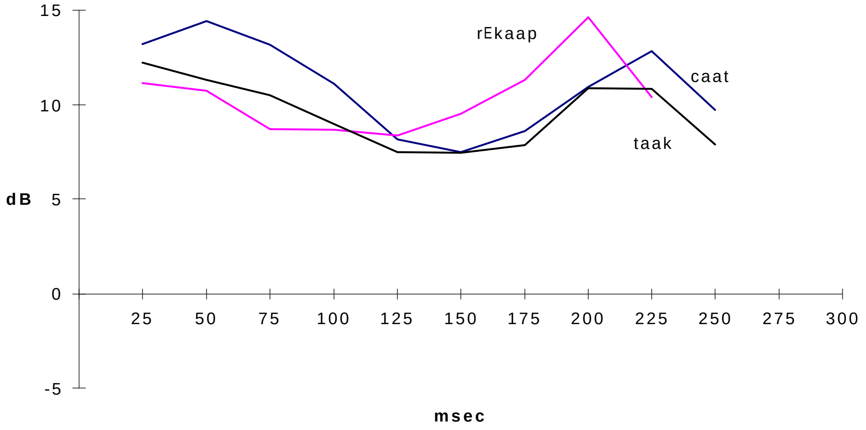


Figure 95.
CHONG, H1-F2, 8 speakers. Breathy vowels.

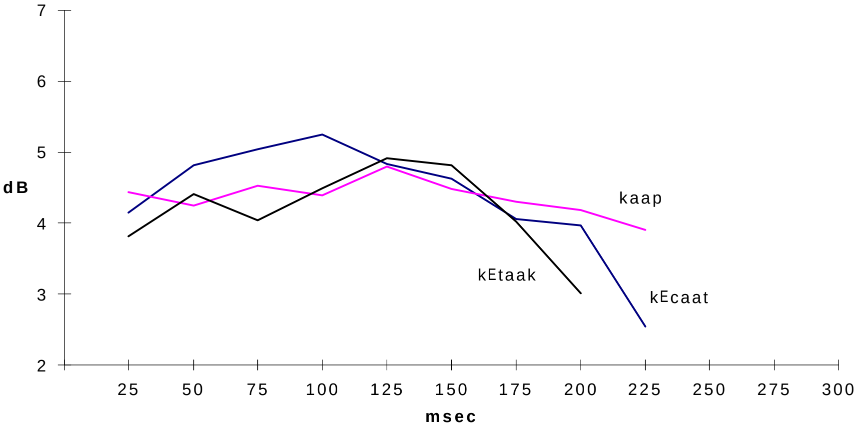


Figure 96.
CHONG, cepstral peak, 8 speakers. Modal vowels.

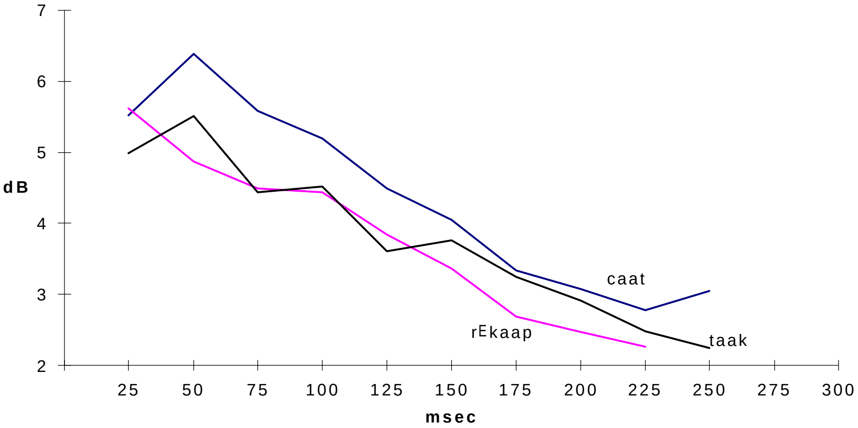


Figure 97.
CHONG, cepstral peak, 8 speakers. Breathy vowels.

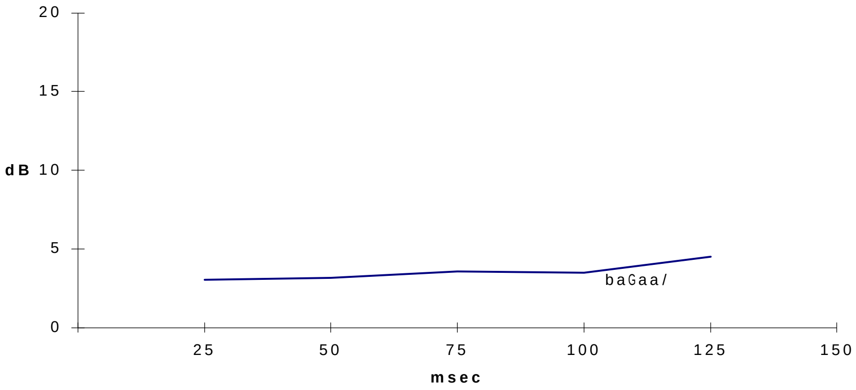


Figure 98.
NAVAJO, H1-H2, 11 speakers. Vowels after stop.

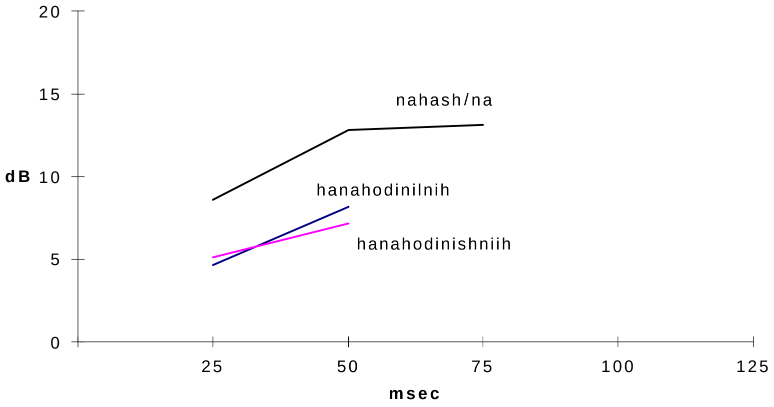


Figure 99.
NAVAJO, H1-H2, 11 speakers. Vowels after [h].

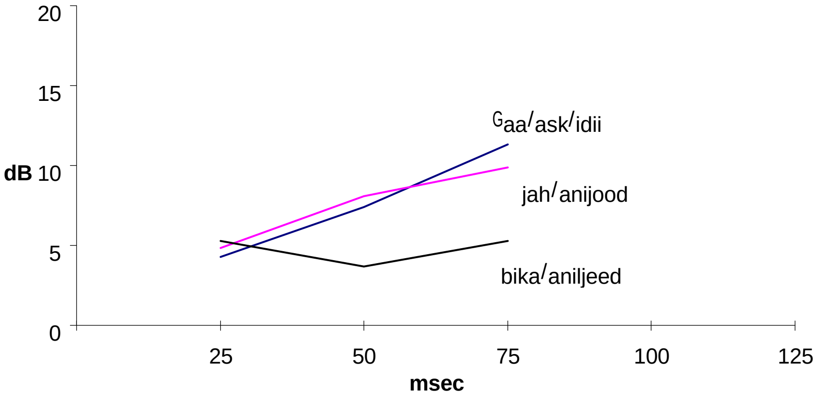


Figure 100.
NAVAJO, H1-H2, 11 speakers. Vowels after [/].

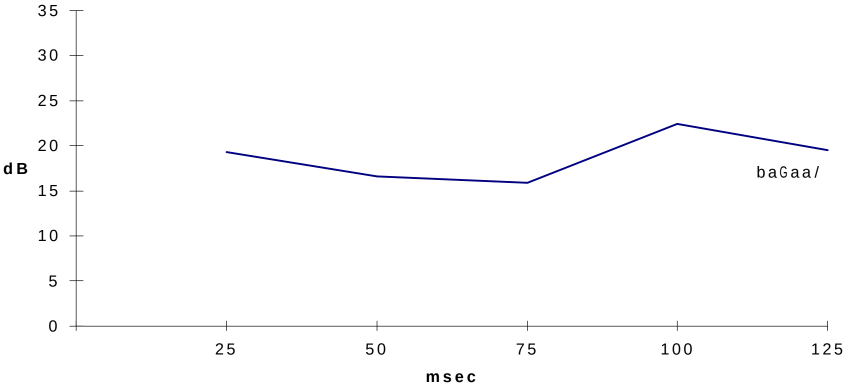


Figure 101.
NAVAJO, H1-F2, 11 speakers. Vowels after stop.

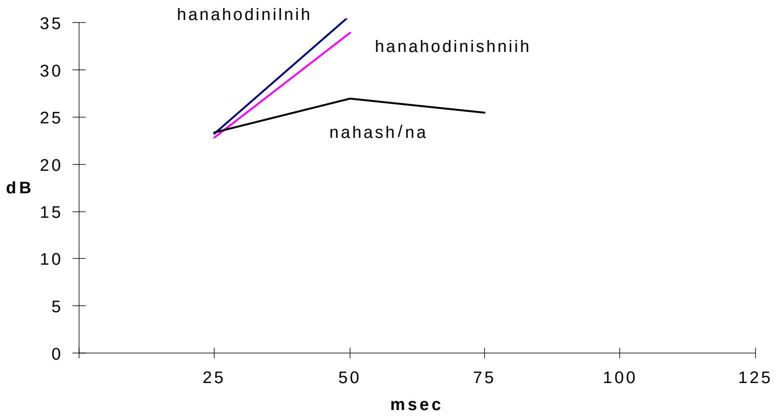


Figure 102.
NAVAJO, H1-F2, 11 speakers. Vowels after [h].

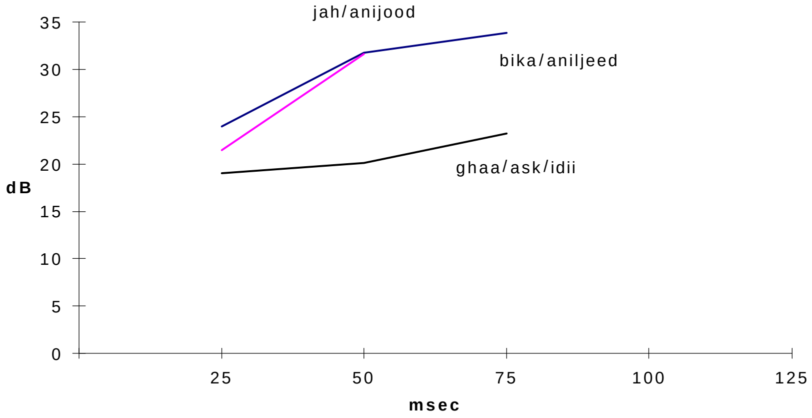


Figure 103.
NAVAJO, H1-F2, 11 speakers. Vowels after [/].

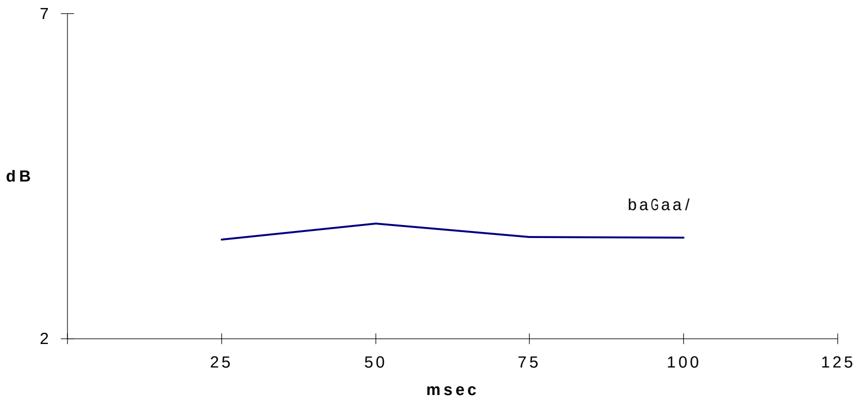


Figure 104.
NAVAJO, cepstral peak, 11 speakers. Vowels after stop.

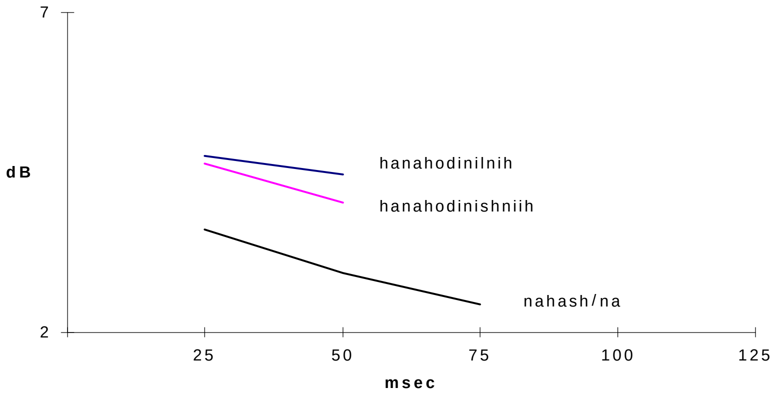


Figure 105.
NAVAJO, cepstral peak, 11 speakers. Vowels after [h].

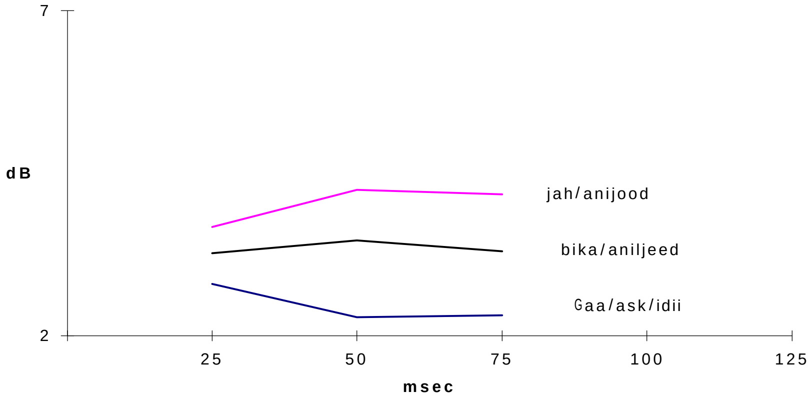
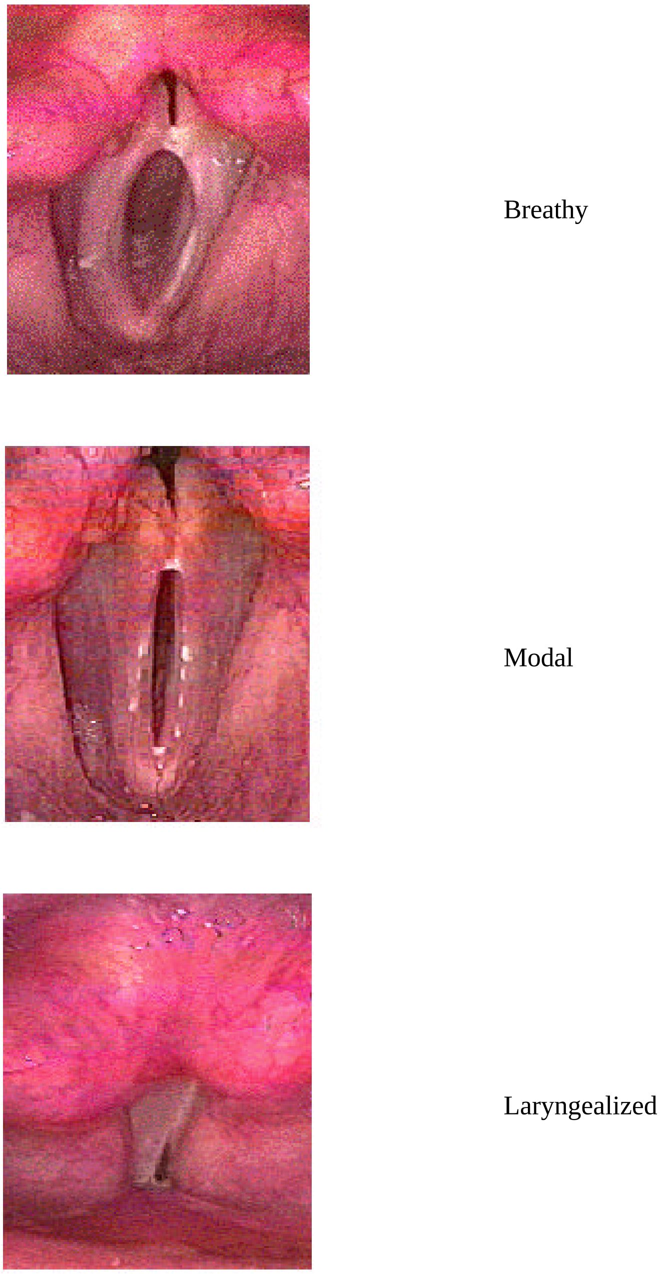


Figure 106.
NAVAJO, cepstral peak, 11 speakers. Vowels after [/].



Breathy

Modal

Laryngealized

Figure 107.
Photographs of a glottis producing three phonation types,
from Ladefoged (2000).

References

Bangayan, Philbert, Christopher Long, Abeer A. Alwan, Jody Kreiman, & Bruce R. Gerratt, 1997. Analysis by synthesis of pathological voices using the Klatt synthesizer. *Speech Communication* **22**, 343-368.

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

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

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

Cho, Taehong (1996). *Vowel correlates to consonant phonation. An acoustic-perceptual study of Korean obstruents*. Master's thesis, University of Texas at Arlington.

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

Fischer-Jorgensen, Eli (1967). Phonetic analysis of breathy (murmured) vowels in Gujarati. *Indian Linguistics* **28**, 71-139.

Gobl, Christer, & Ailbhe Ni Chasaide (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, Christer, & Ailbhe Ni Chasaide (1992). Acoustic characteristics of voice quality. *Speech Communication* **11**, 481-490.

Grimes, Barbara F., ed. (1996). *Ethnologue.* Consulting editors: Richard S. Pittman & Joseph E. Grimes. Dallas, TX: Summer Institute of Linguistics, Inc. http://www.sil.org/ethnologue

Hammarberg, Britta, Bjorn Fritzell, Jan Gauffin, & Johan Sundberg (1986). Acoustic and perceptual analysis of vocal dysfunction. *Journal of Phonetics* **14**, 533-547.

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

Henton, Caroline G., & R. Anthony W. Bladon (1985). Breathiness in normal female speech: inefficiency versus desirability. *Language Communication* **5**, 221-227.

Hillenbrand, James, Ronald A. Cleveland, & Robert L. Erickson (1994). Acoustic correlates of breathy vocal quality. *Journal of Speech and Hearing Research* **37**, 769-778.

Holmberg, Eva B., Robert E. Hillman, Joseph S. Perkell, Peter C. Guiod, & Susan L. Goldman (1995). Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice. *Journal of Speech and Hearing Research* **38**, 1212-1223.

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

Kirk, Paul, Jenny Ladefoged, & Peter Ladefoged (1984). Quantifying acoustic properties of modal, breathy, and creaky vowels in Jalapa Mazatec. *UCLA Working Papers in Phonetics* **59**, 102-113.

Klatt, Dennis H., & Laura C. Klatt (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, Peter (1971). *Preliminaries to linguistic phonetics*. Chicago: The University of Chicago Press.

Ladefoged, Peter (1983). The linguistic use of different phonation types. *Vocal fold physiology. Contemporary research and clinical issues*, edited by Diane M. Bless and James H. Abbs, 351-360. San Diego: College-Hill Press.

Ladefoged, Peter (2000). *Vowels and Consonants*. Oxford: Blackwell Publishers. Accompanying CD with audio and visual illustrations is also available at http://www.humnet.ucla.edu/linguistics/VowelsandConsonants.

Ladefoged, Peter, & Norma Antonanzas-Barroso (1985). Computer measures of breathy phonation. *UCLA Working Papers in Phonetics* **61**: 79-86.

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

Ladefoged, Peter, Ian Maddieson, & Michel Jackson (1988). Investigating phonation types in different languages. *Vocal physiology: voice production, mechanisms, and functions,* edited by Osamu Fujimura, 297-317. New York: Raven Press. Ltd.

Laver, John (1980). *Phonetic description of voice quality*. Cambridge: University of Cambridge Press.

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

Lofqvist, Anders, & Richard S. McGowan (1992). Influence of consonantal envelope on voice source aerodynamics. *Journal of Phonetics* **20**, 93-110.

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

Pierrehumbert, Janet, & David Talkin (1992). Lenition of /h/ and glottal stop. *Papers in laboratory phonology II. Gesture, segment, prosody*, edited by Gerard J. Docherty & D. Robert Ladd, 90-119. Cambridge: Cambridge University Press.

Rabinov, C.R., J. Kreiman, B.R. Gerratt, & S. Bielamowicz (1995). Comparing reliability of perceptual and acoustic measures of voice. *Journal of Speech and Hearing Research* **38**, 26-32.

Schachter, Paul (1987). Tagalog. *The world's major languages,* edited by Bernard Comrie, 933-958. London & Sydney: Croom Helm.

Silverman, Daniel (1995). *Phasing and recoverability*. Doctoral dissertation, University of California at Los Angeles.

Silverman, Daniel, Barbara Blankenship, Paul Kirk, & Peter Ladefoged (1995). Phonetic structures in Jalapa Mazatec. *Anthropological Linguistics* **37**, 70-88

Stevens, Kenneth N. (1977). Physics of laryngeal behavior and larynx modes. *Phonetica* **34**: 264-279.

Stevens, Kenneth N. (1988). Modes of vocal fold vibration based on a two-section model. *Vocal physiology: voice production, mechanisms, and functions,* edited by Osamu Fujimura, 357-371. New York: Raven Press. Ltd.

Stevens, Kenneth N., & Helen M. Hanson (1995). Classification of glottal vibration from acoustic measurements. *Vocal fold physiology. Voice quality control,* edited by Osamu Fujimura & Minoru Hirano, 146-170. San Diego, CA: Singular Publishing Group, Inc.

Titze, Ingo R. (1995). Definitions and nomenclature related to voice quality. *Vocal fold physiology. Voice quality control,* edited by Osamu Fujimura & Minoru Hirano, 335-342. San Diego, CA: Singular Publishing Group, Inc.

Todaka, Yuichi (1993). *A cross-language study of voice quality: bilingual Japanese and American English speakers*. Doctoral dissertation, University of California at Los Angeles.

UCLA Department of Linguistics, Phonetics Laboratory (1991). *Sounds of the world's languages*. Hypercard stacks. Los Angeles: University of California.

Biographical note

During 20 years as an opera singer in southern California and northern Germany, Barbara Blankenship developed an interest in both the sounds of languages and the process of phonation. Doctoral research at the phonetics laboratory of the UCLA Department of Linguistics provided an opportunity to combine the two interests, resulting in this dissertation.

A native of San Diego, Ms. Blankenship received her bachelor's degree in music from Pomona College, and her master's and doctoral degrees in linguistic phonetics from UCLA. She lives in Los Angeles, where she works as a database consultant.