Pitch discrimination during breathy versus modal phonation[*]

Daniel Silverman

Department of Linguistics, University of Illinois

dan@uiuc.edu

X.1 Introduction

The typology of linguistic sound patterning indicates an extreme dispreference for the simultaneous implementation of contrastive tone and contrastive non-modal phonation. Accordingly, I performed a psychoacoustic experiment to investigate whether human auditory perceptual limitations may play a role in this aspect of phonological systems. The experiment consisted of subjects listening to pairs of stimuli—modally phonated pairs, and breathy pairs, deriving from the natural speech of Jalapa Mazatec speakers—which differed in pitch to varying degrees. Subjects were asked to judge whether the two stimuli were the same or different in pitch. I found that, indeed, listeners are better at discerning pitch differences in complex tones implemented during modal phonation than they are discerning pitch differences implemented during breathy phonation.

X.2 Background

Moore (1989) discusses a number of relevant findings in the domain of human perception of complex tones. I summarize here the most pertinent among them.

The lowest harmonics up to the fifth seem to dominate for the purpose of pitch perception. In this harmonic range, resolution of individual overtones is possible. A pattern recognition model of pitch perception (deriving from Terhardt's models, e.g. 1974, and studies by Ritsma 1962, 1963,

[*] This research was supported by NIH Training Grant T32 DC 00008. Thanks to Norma Antoñanzas-Barroso, Bruce Gerratt, and Jody Kreiman for their support at every stage of this study. Thanks also to Khalil Iskarous for technical assistance.

1967, and Plomp 1964) has been offered to account for this finding, in which the values of individual harmonics is determined, and the distance between these harmonics is subsequently calculated, culminating in a pitch percept. An anonymous LabPhon reviewer points out however that just-noticeable differences (JNDs) found in human pitch perception cannot be explained by this theory alone. Auditory filter bandwidths (critical bands) are too wide by a factor of about 30 compared with the JND for pitch, and therefore the pattern recognition model cannot rely solely on the individual harmonics alone. The temporal theory plays a role here since it relates the neural firings to auditory filter outputs, and thus can account for JNDs. At higher frequency levels, where individual harmonics are not resolvable, Moore argues that a temporal model—deriving from the work of Schouton (e.g. 1970)—may best account for the existence of a pitch percept: the pitch percept is ascribed to the temporal rate of repetition of the complex waveform. Moore concludes that an account of pitch perception at both lower and higher frequency ranges is best characterized by a "spectro-temporal" model, in which both harmonic structure and pulse period are relevant.

Moore's report is based on studies in which the experimental stimuli consist of a high signal-to-noise ratio, and well as a periodic rate of repetition, thus mimicking certain qualities of normal human phonation. However, if stimuli were to depart from normal phonation in consisting of a marked lowering of signal-to-noise ratio, as well as a less periodic rate of repetition, it remains to be seen whether subjects are equally adept at perceiving and discriminating pitches.

So-called breathy phonation possesses both these qualities: lower signal-to-noise ratio, and moderate pulse period irregularity. For example in Jalapa Mazatec (an Otomanguean language of Oaxaca, Mexico) breathy vowels involve a marked decrease in signal-to-noise ratio (Silverman, Blankenship, Kirk, & Ladefoged 1995). Moreover, the glottal pulse period in breathy vowels is irregular (Kirk, Ladefoged, & Ladefoged 1993). For these reasons, pitch perception during Jalapa Mazatec breathy phonation may be less accurate than pitch perception during modal phonation.

Briefly now, consider the place of Jalapa Mazatec in the typology of tonal and phonation contrasts.

Some languages are tonal, such as Mandarin Chinese. Here, modally phonated vowels possess tone (although the low tone may occasionally possess a creaky quality).

Table X.1 Mandarin Chinese:

high level	tʰan˥	*greedy*
mid rising	tʰan˧˥	*deep*
low (-rising)	tʰan˩(˩˦)	*perturbed*
high falling	tʰan˥˩	*spy*

In contrast, some languages have contrastive breathiness while lacking tone. Gujarati is one such language (Patel and Mody 1961; Fischer-Jørgensen 1970; Taylor 1985). In such languages, breathy phonation is typically implemented for the duration of the vocalic gesture, and oftentimes into sonorant codas as well.

Table X.2 Gujarati (Fischer-Jørgensen 1970: no glosses provided):

tʃi̤r	mo̤r	dṳd
bi̤	do̤ɽ	pe̤lo
se̤dʒ	ko̤ɽ	ta̤ro
me̤k	ko̤	wa̤li
ba̤r	po̤r	kə̤ɽi

Third, some tone languages possess non-modal phonation contrasts as well as tone. But while a full array of tonal patterns is found on modally phonated vowels, non-modally phonated vowels never contrast for tone. White Hmong exemplifies this pattern (Lyman 1974; Smalley 1976; M.K. Huffman 1987; Ratliff 1992). Breathy phonation here is reportedly implemented for the duration of the vowel.

Table X.3 White Hmong:

High	tau˥	*pumpkin*
Rising	tau˧˥	*to dam up (water)*
Low	tau˩	*axe*
Mid (normal)	tau˧	*to be able*
Falling (normal)	tau˥˧	*sp. of grass*
Creaky	ta̰ṵ˧˩	*bean*
Breathy	ta̤ṳ˥˧	*to follow*

Finally, some languages, such as Jalapa Mazatec, possess vowels in which tone and non-modal phonation fully cross-classify, that is, both tonal and phonatory contrasts may reside on a single vowel. Significantly, vowels which possess both tone and contrastive breathiness here are realized in a part-breathy–part-modal fashion. Specifically, the first portion of the vowel is breathy, while the latter portion is more or less modal. Anticipating my findings now, since pitch (<tone) is more reliably distinguished during modal phonation, a portion of the vowel is given to plain voicing, where tone contrasts are presumably more salient. The remaining portion of the vowel, however, is breathy. In this way, I suggest that both phonation and tone contrasts are effectively conveyed to the listener, with tone residing on the vowel's latter, modal portion.

Table X.4 Jalapa Mazatec:

mm̤æ̤æː˥	*wants*
nn̤a̤a˥	*my tongue*

ɲɲ̤a̤a˧	*nine*
jj̤æ̤æ˩	*boil*
ww̤o̤o˧	*hungry*

Indeed, throughout the Otomanguean family, tone and non-modal phonation are temporally arranged in various ways, even contrastively within a single language, as shown schematically in Table X.5. Note that there is phonological evidence that all the patterns in Table X.5 are treated as monosyllables (see especially Longacre 1952, and Silverman 1995, 1997). Thus Mazatec possesses vowels which may be breathy during their first portion, while contrastive tone values reside on the vowels' latter, modally phonated portion. Chinantec posseses this pattern, but also possesses a contrastive pattern in which breathy phonation *follows* modal phonation. Finally, Trique possesses both these contrastive patterns, and also possesses so-called vocalic "interruption" in which the laryngeal intrudes upon the middle portion of the toned vowel.

Table X.5 Timing contrasts between tone and breathy phonation in Otomanguean

Mazatec	CV̤V́		
Chinantec	CV̤V́	CV́h	
Trique	CV̤V́	CV́h	CV́ɦV́

Indeed, it is this very fact that provides the impetus for the present study. Specifically, why should breathy phonation be limited to only a portion of the vowel in these languages, whereas the canonical realization of breathy vowels in fact involves breathiness throughout? Might there indeed be an auditory incompatibility between tone and breathiness which has lead to this unusual sequencing of articulatory events? This is not to say, of course, that articulatory incompatibilities might not be playing a role here as well. That is, particular phonation-based gestures and pitch-based gestures may be difficult to simultaneously implement, or to implement in rapid succession, as they may make conflicting articulatory demands on the vocal apparatus.

The present study sets out to investigate only the issue of perceptual discriminability, however. Listeners were asked to discriminate pitch values between pairs of stimuli that were either modal or breathy in their phonatory characteristics. The stimuli themselves were derived from actual Jalapa Mazatec vowels. Results indicate that, indeed, subjects are better at distinguishing pitch values implemented during modal phonation than they are distinguishing pitch values implemented during breathy phonation. Implications for cross-linguistic tendencies in phonological structure are discussed as well.

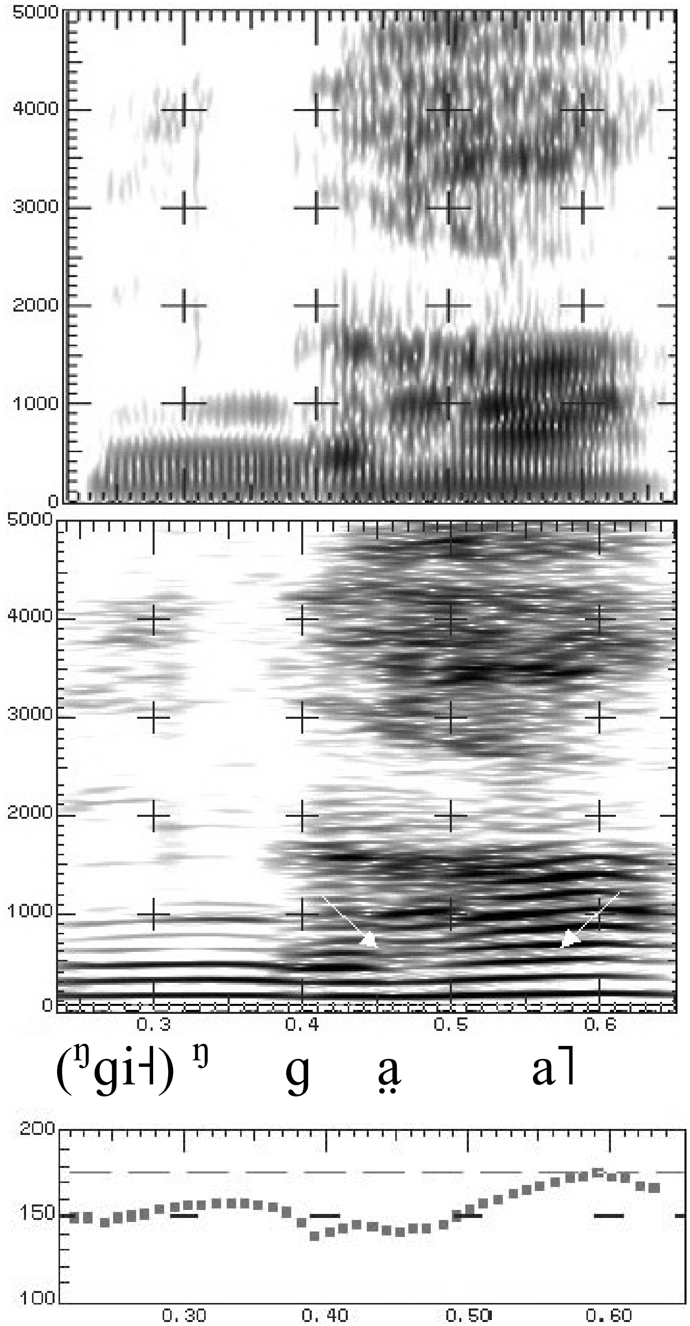
X.3 Stimuli

Three words of Jalapa Mazatec—(a) $^{\text{ŋ}}$gi˧$^{\text{ŋ}}$ga̤a˥ he fastened, (b) $^{\text{n}}$da̤a˥ hard, and (c) mm̤æ̤æ˥ he wants—were digitized with the Kay Computer Speech Laboratory. The first two words were spoken by normal adult male native speakers, while the third word was spoken by a normal adult female native speaker. Recordings were made in an outdoor setting in the village of Jalapa de Diaz, Oaxaca, Mexico, by Keith Johnson, Paul Kirk, Peter Ladefoged, Joyce McDonough, and Donca Steriade, employing a Marantz portable cassette recorder which had a frequency response that was better than ±2 dB over the range 70 - 10,000 Hz, and had a signal-to-noise ratio of 42 dB.

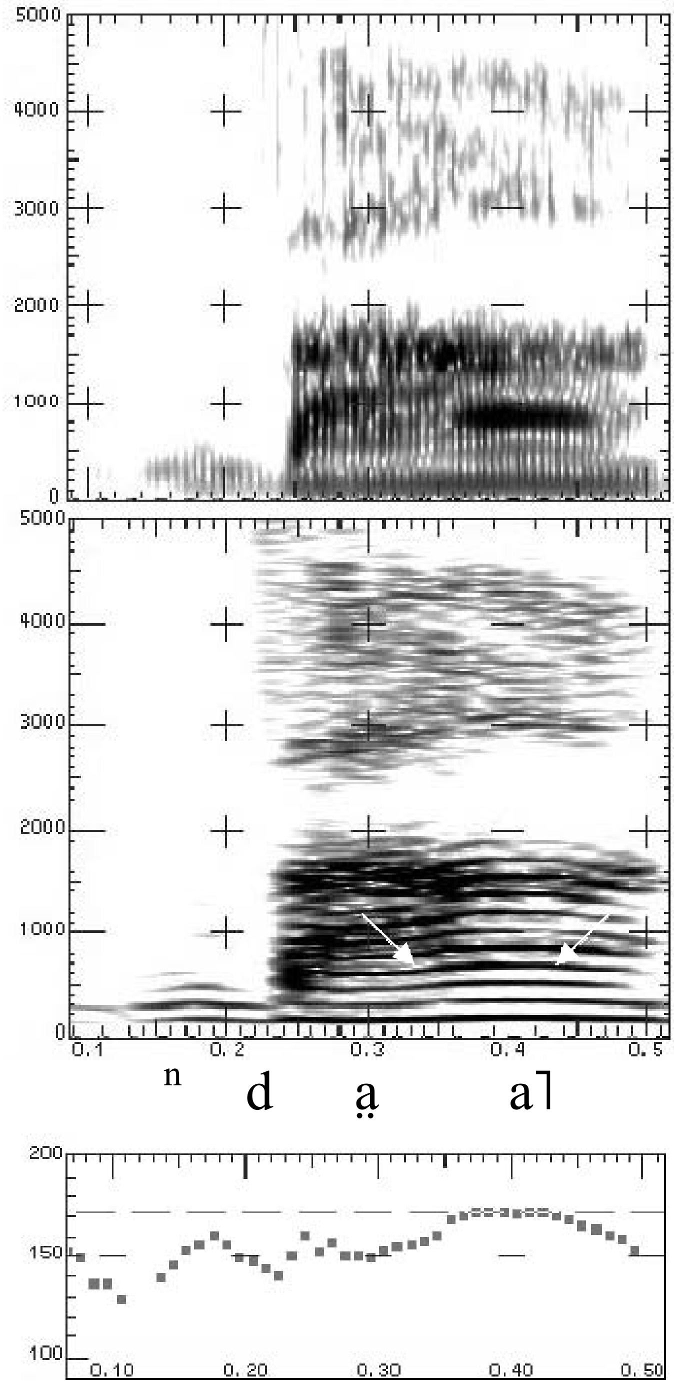
Wideband and narrowband spectrograms of these source data, as well as their pitch tracks, are in Figure 1. Note in particular the shift in phonation during the vowel. This is best seen in the narrowband spectrograms, in which noise accompanies the harmonic structure during the early, breathy portion of the vowel, though is significantly reduced during the latter, modal portion. Note especially the fourth harmonic, indicated with arrows, which is within what might be the most important region for the perception of pitch during speech. The third form, mm̤æ̤æ˥, possesses a significantly less marked transition from breathy to modal phonation. This is likely due to the absence of a stop release in the transition from consonant to vowel here: oral stop releases, as opposed to sonorant releases, are aerodynamically better suited to induce a robust realization of contrastive laryngeal states (Kingston 1985). Also note that pitch height is typically most stable on the modal portion of the vowel. Considering the results of Rosenberg (1965), who finds that when a pulse train varies, or jitters, by more than 10%, an otherwise just-noticeable pitch difference within the 300-1000 Hz range is rendered indiscriminible, the stable realization—and/or (in the case of contours) smooth gliding realization—of pitch is consistent with the hypothesis that tone is better conveyed during modal phonation than during breathy phonation.

Figure X.1 Wideband and narrowband spectrograms, and pitch tracks of digitized speech from Jalapa Mazatec:

(a)

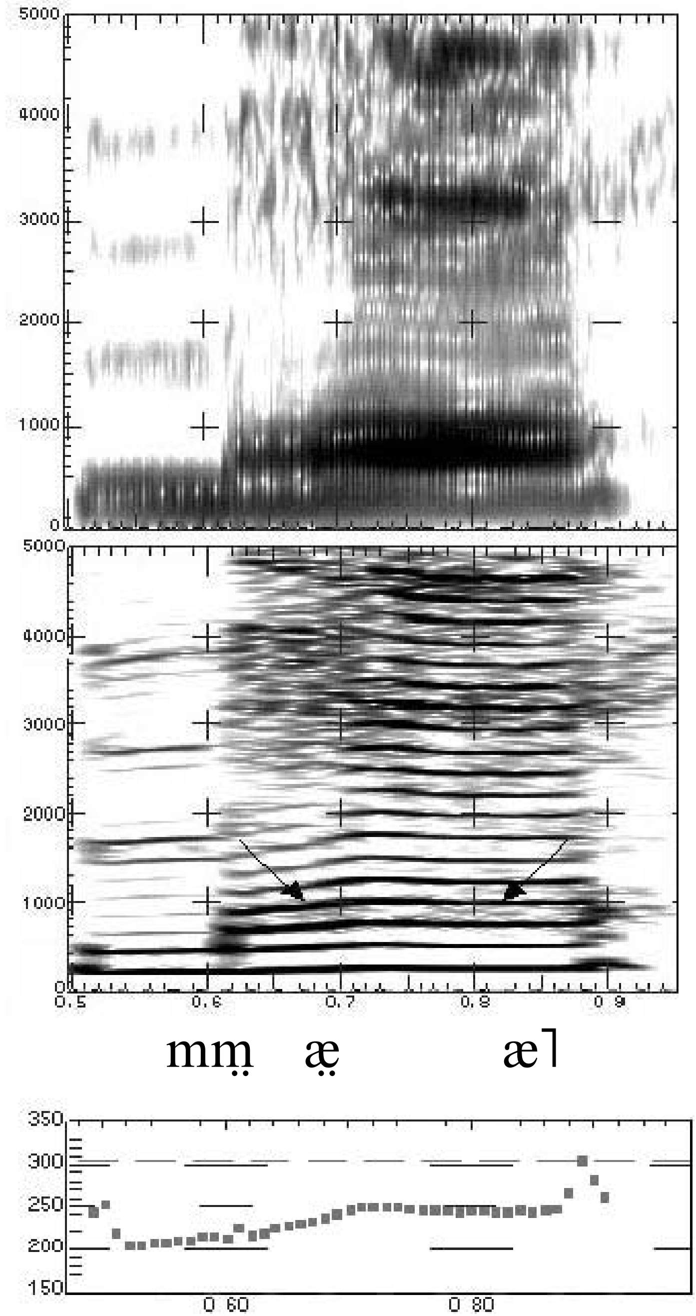


(b)



5000
4000
3000
2000
1000
0
0.1
0.2
0.3
0.4
0.5
n
d
a̤
a˥
200
150
100
0.10
0.20
0.30
0.40
0.50

(c)

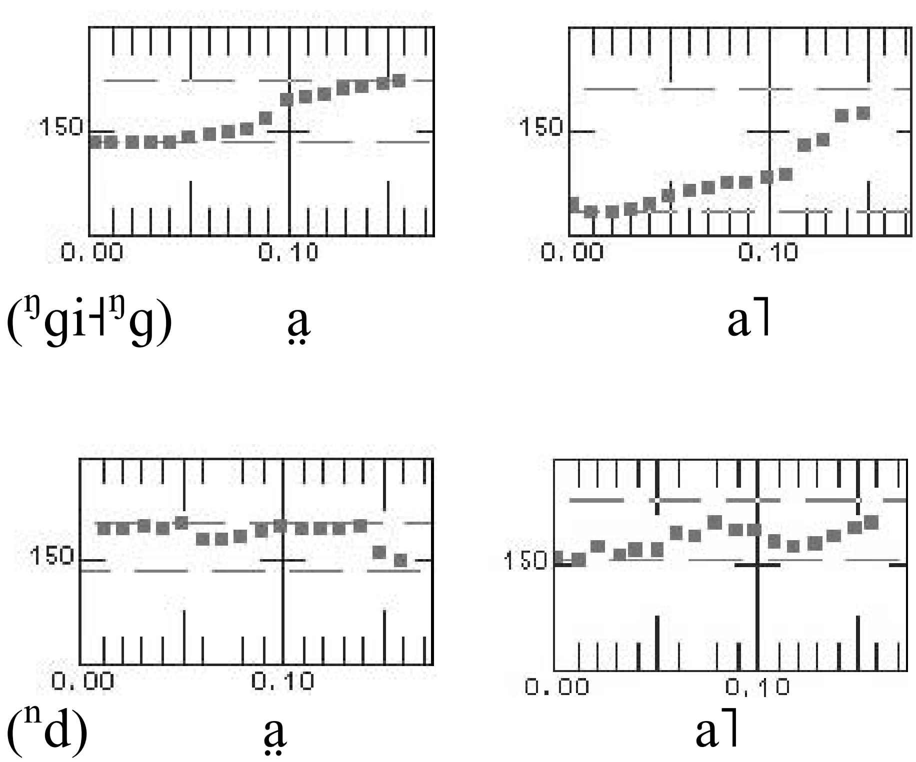

5000
4000
3000
2000
1000
0
0.5
0.6
0.7
0.8
0.9
mm̤ æ̤ æl
350
300
250
200
150
0 60
0 80

After digitization, both the breathy portion and modal portion of each word were extracted. As the breathy component of Jalapa Mazatec vowels is lower in pitch than its associated modal component in the case of high tones, I lowered the pitch of modal portions to approximate the pitch of breathy portions, employing the SoundEdit16.2 "bender" feature. Pitch tracks for the base stimuli are in Figure X.2.

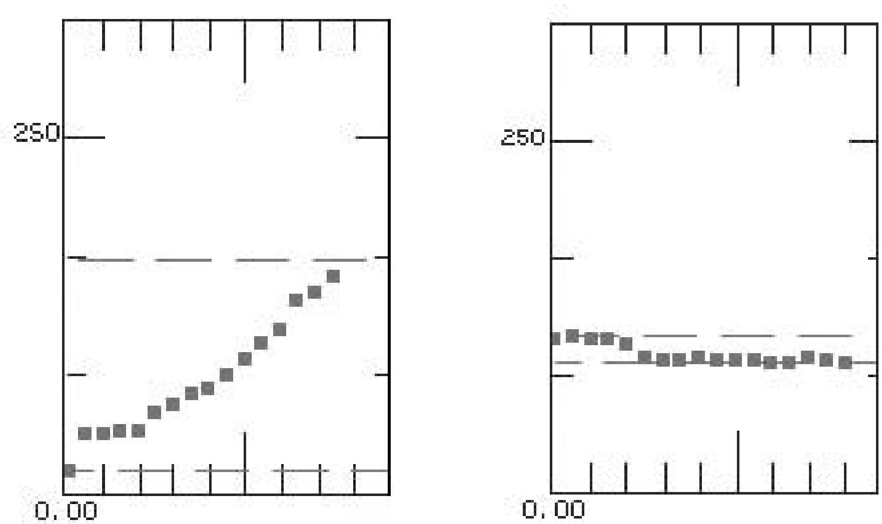
Figure X.2 Pitch tracks for base stimuli

(a)

(b)



(c)



(mm̤) æ̤ æ˥

The peak amplitudes of the six waveforms of the stimuli were normalized, and onsets and offsets were ramped in order to avoid click artifacts. The respective fundamental frequencies of each waveform was increased in increments of approximately 3 Hz, up to 24 Hz, which resulted in six continua with nine steps each.

For the pitch shifts, the signal was simply sped up. The playback sample rate was manipulated and the sound resampled to the original sample rate. In this procedure, spectra are shifted in frequency and thus the ratios of component frequencies are preserved. Given the spectral shift involved, some slope distortion may be added to the modified signal: formants are shifted downwards in slowed-down forms, upwards in sped-up forms. But given the very minor signal adjustments employed in this study, spectral shifts are small, and, especially important, equally present in both the modal and breathy continua. To exemplify this effect, approximate F1 values for the baseline, fourth, and final steps for the ^{n}dạa˥ continua are presented in Table X.6.

Table X.6 F1 values for baseline, fourth, and final steps for the ⁿda̤a˥ continua.

	F1, modal [a]	F1, breathy [a̤]
Baseline:	841 Hz	979 Hz
Fourth step:	879	1141
Final step:	935	1187

Within-continuum increases in F1 values are a consequence of speeding up the recording, while the lower F1 values present in the modal continuum are a consequence of slowing down the baseline form to approximate the pitch of the breathy baseline form.

All forms were then converted to 200msec in length. Every form was paired with every other form within its continuum, unless the given pair would exceed 12 Hz difference. This resulted in a total of 366 stimulus pairs, as schematized in Table X.7. Thus, for example, one form possessing a fundamental frequency 21 Hz above the baseline was paired with all forms between 9 and 24 Hz above the baseline: forms between the baseline and 9 Hz above the baseline were not paired with this form, however, as their differences exceed 12 Hz.

Table X.7 Schematic of all paired stimuli for one of the six continua

	0 Hz	3 Hz	6 Hz	9 Hz	12 Hz	15 Hz	18 Hz	21 Hz	24 Hz
0 Hz									
3 Hz									
6 Hz									
9 Hz									
12 Hz									
15 Hz									
18 Hz									
21 Hz									
24 Hz									

X.4 Subjects and procedure

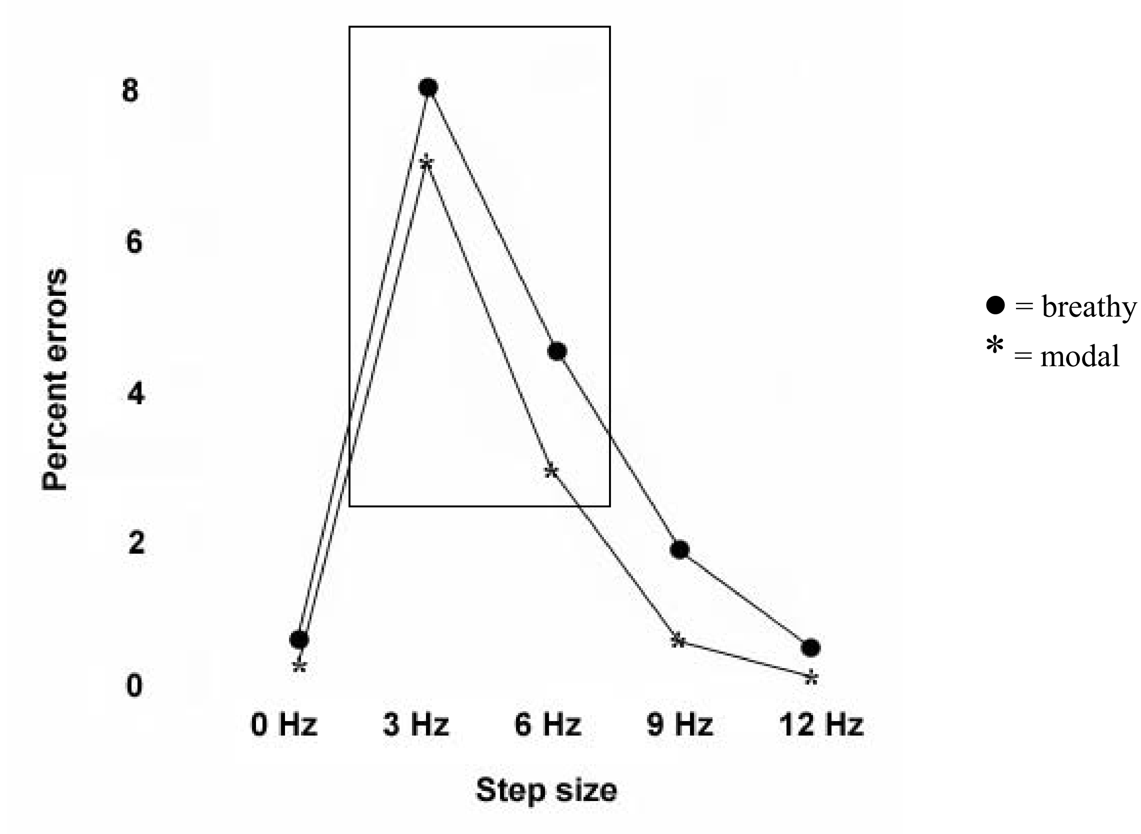
10 non-Jalapa Mazatec-speaking UCLA graduate students in phonetics/phonology listened (individually) in a sound booth to 1000 trials each (501 “different” pairs; 499 “same” pairs), presented in blocks of 50. The inter-stimulus interval was 300 msec, while the inter-trial interval was 3 sec. Subjects were asked to judge for each pair whether the two stimuli were the same or different in pitch.

Non-Jalapa Mazatec speakers were chosen because they are readily available, but also because they might be more capable of detecting the exceedingly minor pitch differences involved, as they are less likely to be influenced by linguistically-based tonal categorical perception.

X.5 Results

Subjects performed more accurately on modal vowel pairs than on breathy vowel pairs (ANOVA, $p<.05$). This is indicated by the higher error rate in breathy pairs versus corresponding modal pairs in Figure X.3. Moreover, at the 3 Hz and 6 Hz intervals, performance was significantly worse than performance at the 9- and 12-Hz intervals (Scheffé, $<.05$; see the boxed values in Figure X.3).

Figure X.3 Results



Thus, not only was subject performance significantly worse overall on breathy token pairs, but also, subjects performed significantly worse as the pitch interval between tokens fell to approx. 6 Hz and below.

In the debriefing interview, most subjects reported being unaware that the stimuli were derived from actual language, and instead assumed that they were computer-generated “bleeps,” suggesting that linguistically-based categorical perception would not have been a factor, regardless of the native language of the subjects.

X.6 Discussion and conclusion

The results of this study may be seen as complementing those of Rosenberg (1965), who, recall, found that when a pulse train varies, or jitters, by more than 10%, an otherwise just-noticeable pitch difference within the 300-1000 Hz range is rendered indiscriminible. Thus whether jittered (a common acoustic correlate of vocalic “creak”) or reduced in signal-to noise ratio (a common acoustic correlate of vocalic breathiness), or perhaps especially both, pitch perception during non-modal phonation suffers.

Of course, experimental data cannot be generalized directly to natural linguistic data. However, the results of the present study suggest that tonal and phonation contrasts have the distributions they do for good reason. Specifically, although it is only in an experimental setting, as opposed to a natural linguistic setting, that listeners may be called upon to determine just- and near-just-noticeable differences in pitch, it should not be surprising that languages might evolve to avoid less-good contrasts in favor of better ones. Such a “sensitive dependence on initial conditions” (Gleick 1987:8), is fully consistent with the hypothesis that minor phonetic distinctions that are never employed in phonological systems might nonetheless constitute the “phylogenetic” origin of phonetic distinctions that *are* linguistically relevant.

There are in fact various tendencies in phonological systems that support this line of reasoning. First, trained subjects are able to discriminate minor differences in voice onset time (VOT) that are never employed contrastively in language (see, for example, Strange 1972). Languages typically employ VOT differences that are far less effortfully noticeable; positive VOT (aspirated), zero VOT (plain), and negative VOT (voiced).

Second, nasal vowels tend to possess fewer quality contrasts than oral vowels do. For example, many American English dialects have lost the “pin-pen” contrast. The standard account of this asymmetry implicates the presence of the nasal pole and zero structure, which is superimposed on the oral formant structure. This superimposition clearly does not obliterate the pin-pen distinction in all dialects, but might nonetheless make it less likely that a language should exploit the oral space as fully in this context, thus possibly influencing this diachronic merger.

Third, certain low-level phonetic information may come to take a prominent phonological role as other cues undergo diachronic attrition. The pronounced English vowel length distinction exemplifies this phenomenon, in which a difficult voicing contrast in coda stops has been displaced on to a primary vowel length contrast (for example, Raphael 1971). Similarly, certain cases of “tonogenesis,” in which another difficult voicing contrast evolves into a tonal one (as in Cantonese), shows that rather insignificant automatic features can come to play prominent functional roles in the system of contrasts.

While the discussed implications of the present findings—that distinctions which phonological systems never exploit might nonetheless constrain phonological patterning at a phylogenetic distance—are more speculative than both the second and third cases just discussed, they nonetheless should not be dismissed out of hand. To the extent that parallels can be observed between natural language typology and experimentally ascertained perceptual asymmetries, psychoacoustic experimentation along the present lines may constitute a potentially fruitful base for phonological theorization.

References

Fischer-Jørgensen, E. (1970) “Phonetic analyses of breathy (murmured) vowels in Gujarati,” Indian Linguistics 28:71-140.

Gleick, J. (1987) Chaos. New York, Penguin.

Kirk, P.L., J. Ladefoged, & P. Ladefoged (1993) “Quantifying acoustic properties of modal, breathy, and creaky vowels in Jalapa Mazatec,” in A. Mattina and T. Montler, eds., American Indian Linguistics and Ethnography in Honor of Lawrence C. Thompson. Occasional Papers in Linguistics 10, University of Michigan, 435-450.

Huffman, M.K. (1987) “Measures of phonation types in Hmong,” Journal of the Acoustical Society of America 81.2:495-504.

Kingston, J. (1985) “The phonetics and phonology of the timing of oral and glottal events,” Ph.D. dissertation, University of California at Berkeley.

Longacre, R. (1952) “Five phonemic pitch levels in Trique,” Acta Linguistica VII,1-2:62-82.

Lyman, T. (1974) Dictionary of Hmong Njua. The Hague, Mouton.

Moore, B.C.J. (1989) An introduction to the psychology of hearing, 3rd edition. London, Academic.

Patel, M.S. & J.J. Mody (1961) The vowel system of Gujarati. Faculty of Education and Psychology, Maharaja Sayajirao University of Baroda, Baroda.

Plomp, R. (1967) “Pitch of complex tones,” Journal of the Acoustical Society of America 41:1526-1533.

Raphael, L.J. (1971) “Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English,” JASA 51:1296-1303.

Ratliff, M. (1992) Meaningful tone: a study of tonal morphology in compounds, form classes, and expressive phrases in White Hmong. Center for Southeast Asian Studies, Monograph series on Southeast Asia, Special Report 27, Northern Illinois University.

Ritsma, R.J. (1962) “Existence region of the tonal residue I,” Journal of the Acoustical Society of America 34:1224-1229.

Ritsma, R.J. (1963) “Existence region of the tonal residue II,” Journal of the Acoustical Society of America 35:1241-1245.

Ritsma, R.J. (1967) “Periodicity detection,” in R. Plomp and G.F. Smoorenburg, eds., Frequency analysis and periodicity detection in hearing. Leiden, Sijthoff.

Rosenberg, A.E. (1965) “Pitch discrimination of jittered pulse trains,” Journal of the Acoustical Society of America 39.5:920-928.

Schouton (1970) “The residue revisited,” in R. Plomp and G.F. Smoorenburg, eds., Frequency analysis and periodicity detection in hearing. Leiden, Sijthoff.

Silverman, D. (1995) Phasing and recoverability. UCLA dissertation. Published 1997 in Outstanding Dissertations in Linguistics series. New York, Garland.

Silverman (1997) “Laryngeal complexity in Otomanguean vowels,” Phonology 14.2.

Silverman, D., B. Blankenship, P. Kirk, & P. Ladefoged (1995) “Phonetic structures in Jalapa Mazatec,” Anthropological Linguistics 37.1:70-88.

Smalley, W.A. (1976) “The problems of consonants and tone: Hmong (Meo, Miao),” in W.A Smalley, ed., Phonemes and orthography: language planning in ten minority languages in Thailand. Pacific Linguistics Series C, No. 43.

Strange, W. (1972) The effects of training on the perception of synthetic speech sounds: voice onset time. Ph.D. dissertation, University of Minnesota.

Taylor, G.P. (1985) The student's grammar of Gujarati. Asian Educational Services, New Delhi.

Terhardt E. (1974) Pitch, consonance, and harmony,” Journal of the Acoustical Society of America 55:1061-1069.