Pitch discrimination during breathy versus modal phonation

Daniel Silverman
Department of Linguistics, University of Illinois
Hypothesis

- Listeners are better at discriminating pitches implemented during modal phonation than pitches implemented during breathy phonation

Motivation for hypothesis

- Pitch is primarily determined by glottal pulse period and harmonic structure
- Glottal pulse rate in breathy vowels is irregular in Jalapa Mazatec (an Otomanguean language of Oaxaca, Mexico; Kirk, Ladefoged and Ladefoged 1993); spectrum of Jalapa Mazatec breathy vowels involves significant

- Pitch differences may be less reliably discriminable during breathy phonation than during modal phonation
- Linguistic evidence: Tone and breathy phonation are very rarely implemented simultaneously.
● **Mandarin tones**
  high level \( \text{th} \text{an}\| \) greedy
  mid rising \( \text{th} \text{an}\| \) deep
  dipping \( \text{th} \text{an} \| \) perturbed
  high falling \( \text{th} \text{an}\| \) spy
  toneless \( \text{lo} \) (aspect)

● **Gujarati breathy vowels**
  \( \text{tʃɪr} \) \( \text{mɔr} \) \( \text{dud} \)
  \( \text{bɪ} \) \( \text{dɔr} \) \( \text{peolo} \)
  \( \text{sɛdʒ} \) \( \text{kɔr} \) \( \text{taro} \)
  \( \text{mek} \) \( \text{kɔ} \) \( \text{wali} \)
  \( \text{bər} \) \( \text{pɔr} \) \( \text{kəɾi} \)
**White Hmong “tones”**

<table>
<thead>
<tr>
<th>Tone Description</th>
<th>Symbol</th>
<th>Example Word</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>τu tighten</td>
<td>pumpkin</td>
</tr>
<tr>
<td>Rising</td>
<td>τu tighten</td>
<td>to dam up (water)</td>
</tr>
<tr>
<td>Low</td>
<td>τu−</td>
<td>axe</td>
</tr>
<tr>
<td>Mid (normal)</td>
<td>τu−</td>
<td>to be able</td>
</tr>
<tr>
<td>Falling (normal)</td>
<td>τuɐ</td>
<td>sp. of grass</td>
</tr>
<tr>
<td>Creaky</td>
<td>τuɐ</td>
<td>bean</td>
</tr>
<tr>
<td>Breathy</td>
<td>τuɐ</td>
<td>to follow</td>
</tr>
</tbody>
</table>
Jalapa Mazatec

<table>
<thead>
<tr>
<th>mmææ:</th>
<th>wants</th>
</tr>
</thead>
<tbody>
<tr>
<td>nηαα</td>
<td>my tongue</td>
</tr>
<tr>
<td>ηηαα-</td>
<td>nine</td>
</tr>
<tr>
<td>jjææ</td>
<td>boil</td>
</tr>
<tr>
<td>wooo-</td>
<td>hungry</td>
</tr>
</tbody>
</table>

Stimuli

- Digitized natural speech stimuli from Jalapa Mazatec:
ŋgi-ŋgaaŋ (he fastened)
ndazaar (hard)
Both breathy portion and modal portion extracted from each word

(he wants)
• Pitch of modal portions lowered to equal pitch of breathy portions (with SoundEdit16.2 "bender" feature)
• Amplitude of six spectra normalized for peak amplitude
• Onsets and offsets ramped to avoid click artifacts
• Frequency of each portion increased in increments of approximately 3 Hz., resulting in six continua with five steps each.
• All forms converted to 200 msec in length
• All possible within-continuum pairs (up to one-half tone differences) produced, for a total of 366 stimulus pairs
1000 trials/listener (501 “different” pairs; 499 “same” pairs), presented in blocks of 50 pairs. Inter-stimulus interval = 300 msec; inter-trial interval = 3 sec.

Subjects judged for each pair whether the two stimuli were the same or different pitch.

Results
Discussion

- Subjects performed more accurately on modal vowel pairs than on breathy vowel pairs (p<.05).
- Moreover, at the 3- and 6-Hz. intervals, the difference between performances on modal pairs versus breathy pairs was significantly greater than the difference between modal pairs versus breathy pairs at the 9- and 12-Hz intervals (p<.05) (see “…” in graph above).
- Thus, although frequency was increased linearly in both phonation groups, a nonlinear discriminability relationship existed within
each group; not only was subject performance significantly worse overall on breathy token pairs, but also, subjects performed increasingly worse as the pitch interval between breathy tokens fell to approx. 6 Hz. and below, while a similar fall-off in performance was not found in modal pairs at the same pitch interval and below.

The results of this study may be seen as complementing those of Rosenberg (1966), who found that when a pulse period 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 (‘creaky’) or breathy, pitch
perception during non-modal phonation suffers.

These findings may be viewed as consistent with certain typological linguistic facts:

- Some tonal languages possess non-modal phonation contrasts on vowels (e.g. Burmese, White Hmong, and Vietnamese). While a full array of tonal patterns is found on modally phonated vowels, non-modally phonated vowels never contrast for tone. As pitch (<tone) is more reliably distinguished during modal phonation, it is less likely for non-modal phonated vowels to bear contrastive tone.
- Otomanguean languages (such as Jalapa Mazatec) possess vowels in which tone and non-modal phonation fully cross-classify. However, such vowels possess a part-modal – part-breathy or -creaky realization (cf. examples above). As 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 (or creaky).

- Of course, experimental data cannot be generalized directly to natural linguistic data.
However, these results suggest that tonal and phonation contrasts have the distributions they do for good reason.

- More 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.

- That is, phonetic distinctions that are never employed in phonological systems might nonetheless constitute the ‘phylogenetic’
origin of phonetic distinctions that are linguistically relevant. Non-linguistic phonetic experimentation may thus serve as a jumping-off point for this potentially fruitful area of theorization.

A note on the SoundEdit16.2 “bender” feature

The SE16.2 “bender” slows down or speeds up the playback of a sound. The playback sample rate is manipulated and the sound is resampled to the original (and constant) sample rate. The spectra are equally shifted in frequency and thus the ratios of the component frequencies are preserved. Given the spectral shift involved, some slope distortion may be added to the modified signal: a shift up in formants for sped-up playback, and a shift
down for slowed-down playback. But given the very minor signal adjustments employed in this study (roughly 3 Hz. per step), spectral shifts are exceedingly minor, increasing, of course, as more steps are made.
References


Silverman (in press) "Laryngeal complexity in Otomanguean vowels. Phonology 14.


This research was supported by NIH Training Grant T32 DC 00008. Thanks to Norma Antoñanzas-Barroso, Bruce Gerrett, and Jody Kreiman for their support at every stage of this study.