Linguistic voice quality

Pat Keating
UCLA Linguistics Department
Phonation

- **Phonation**: sound production in the larynx, usually by vocal fold vibration (voice, or voicing)

- How fast the folds vibrate determines voice *pitch*; how they move determines voice *quality*

- These vary *across* speakers (people’s voices sound different) and *within* speakers (individuals can adjust vibration)

Ladefoged gif: http://www.linguistics.ucla.edu/faciliti/demos/vocalfolds/vocalfolds.htm
Some examples by John Laver
- 3 major phonation types

- Laver modal voice
- Laver breathy voice
- Laver creaky voice
Phonation types and glottal opening

How large is the glottal opening?

2 vocal folds

back

glottis (space between)

front

Ladefoged’s glottal continuum

IPA diacritics:  \( \tilde{a} \)  \( \tilde{\text{a}} \)

On the breathy side of modal: lax, slack, or lenis
On the creaky side of modal: tense, stiff, fortis, or pressed
Phonation contrasts in languages of the world

- Many languages contrast phonations on vowels and/or consonants
- Common especially in SE Asia, the Americas, India
How do phonation types (on vowels) differ within and across languages?

This talk:
- Cross-language comparison of vowel phonation acoustics: What is the overall phonetic space for vowel voice quality?
- Phonation in tone languages: How do pitch and phonation interact?
Our project: 10 languages from four language families

<table>
<thead>
<tr>
<th>Sino-Tibetan</th>
<th>Hmong-Mien</th>
<th>Oto-Manguean</th>
</tr>
</thead>
<tbody>
<tr>
<td>*Yi (Southern: Xinping &amp; Jiangcheng)</td>
<td>*Hmong (White Hmong)</td>
<td>Mazatec (Jalapa de Diaz)</td>
</tr>
<tr>
<td>• lax vs. tense</td>
<td>• modal vs breathy H-falling tone, creaky L tone, others modal</td>
<td>• breathy vs. modal vs. laryngealized (creaky)</td>
</tr>
<tr>
<td>• crossed with L, M lexical tones</td>
<td>• Black Miao (Shidong Kou)</td>
<td>• fully crossed with lexical tones</td>
</tr>
<tr>
<td>*Bo (Shizong &amp; Xingfucun) – like Yi</td>
<td>• modal vs breathy M tone, creaky L tone, pressed H tone</td>
<td>*Valley Zapotec (Santiago Matatlán and San Juan Guelavia combined)</td>
</tr>
<tr>
<td>*Hani (Luchun) – like Yi</td>
<td></td>
<td>• Modal H tone, creaky H-falling tone, breathy L-falling tone</td>
</tr>
<tr>
<td>*Mandarin (Beijing) – creaky tone3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* 8 languages with electroglossoigraphy
Black Miao fieldwork in Guizhou (Jianjing Kuang)
Yi-languages fieldwork in Yunnan (Jianjing Kuang)
Hmong fieldwork in Minnesota (Christina Esposito)
In Los Angeles

- Mandarin and English students at UCLA
- Gujarati students at USC
- Zapotec speakers in Koreatown

- Mazatec recordings from online UCLA Phonetic Archive
Sample tokens 4 languages
(1 female speaker each language)

<table>
<thead>
<tr>
<th></th>
<th>Breathy</th>
<th>Lax</th>
<th>Modal</th>
<th>Tense</th>
<th>Creaky</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gujarati</td>
<td>ɓaːɾ</td>
<td>--</td>
<td>ɓaɾ</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Hmong</td>
<td>pɔ̰₄²</td>
<td>--</td>
<td>pɔ̰₅²</td>
<td>--</td>
<td>pɔ̱₂¹</td>
</tr>
<tr>
<td>Mazatec</td>
<td>ɓa³⁴</td>
<td>--</td>
<td>ɓa³²</td>
<td>--</td>
<td>ɓa³</td>
</tr>
<tr>
<td>S. Yi</td>
<td>--</td>
<td>ɓe³³</td>
<td>--</td>
<td>ɓe³³</td>
<td>--</td>
</tr>
</tbody>
</table>
New tools for voice analysis

- For acoustic analysis: VoiceSauce
- For physiological analysis: EggWorks, used with VoiceSauce
- Both = UCLA free software

Shue 2010, Shue et al. 2011, Tehrani 2012
VoiceSauce measures, and those used here

- **F0** from STRAIGHT, Snack, or Praat
- **H1**, **H2**, **H4**
- **H2kHz**, **H5kHz**
- **F1-F4** and **B1-B4** from Snack or Praat
- **A1**, **A2**, **A3**

- All * harmonic measures come both corrected (*) and uncorrected for formants
- **H1*-H2**
- **H1*-A1**
- **H1*-A2**
- **H1*-A3**
- **H2*-H4**
- **H4*-H2kHz**, **H2kHz*-H5kHz**
- **Energy**
- **Subharmonic to Harm. Ratio**
- **Cepstral Peak Prominence**
- **Harmonic to Noise Ratios** (4 freq. bands)
- **Strength of Excitation**
Acoustic measures based on harmonics in spectrum
H1-H2 example: Jalapa Mazatec

<table>
<thead>
<tr>
<th>Breathy</th>
<th>Modal</th>
<th>Creaky</th>
</tr>
</thead>
<tbody>
<tr>
<td>ba\textsuperscript{34}</td>
<td>ba\textsuperscript{32}</td>
<td>ba\textsuperscript{3}</td>
</tr>
</tbody>
</table>

Kirk et al. 1984, Garellek & Keating 2011
Acoustic space for phonation across languages

Many acoustic measures of 24 phonation categories from 10 languages, all speakers
24 categories (for non-high, oral vowels after unaspirated consonants, at mid-vowel)

- Bo
  - Lax, Tense
- English
  - Modal
- Gujarati
  - Breathy, Modal
- Hani (Luchun)
  - Lax, Tense
- Hmong
  - Breathy, Modal, Creaky
- Mandarin
  - Modal, Creaky
- Mazatec
  - Breathy, Modal, Creaky
- Miao (Black)
  - Breathy, Modal, Tense, Creaky
- Yi (Southern)
  - Lax, Tense
- Zapotec (Valley)
  - Breathy, Modal, Creaky
Category means

- Each of 17 measures standardized for each speaker
- Mean for each measure for each of 24 phonation categories – across all tokens and speakers
- $= 17 \times 24$ mean measures
Multi-Dimensional Scaling

- MDS is a reduction of high-dimensional data to a low-dimension map of distances that can be visualized.
- Usually used with perception data, but here applied to acoustic data. Each acoustic measure is a dimension, and each category mean has a multi-dimensional physical acoustic distance from all other category means.
- Can test for strength of contribution of measures to dimensions.
2-D acoustic space from MDS

Languages:
- Bo
- English
- Gujarati
- Luchun Hani
- Hmong
- Mandarin
- Mazatec
- Miao (Black)
- Yi (Southern)
- Zapotec (Valley)

Keating et al. 2012
2-D acoustic space from MDS

- Bo
- English
- Gujarati
- Luchun Hani
- Hmong
- Mandarin
- Mazatec
- Miao (Black)
- Yi (Southern)
- Zapotec (Valley)

Keating et al. 2012
2-D acoustic space from MDS

Bo
English
Gujarati
Luchun Hani
Hmong
Mandarin
Mazatec
Miao (Black)
Yi (Southern)
Zapotec (Valley)
Summary, contrast space

- The acoustic-phonetic space for (vowel) voice quality contrasts is largely 2-D: modal-ness vs. glottal aperture
- Both derived from spectral measures (low and low-mid frequencies)
- Each phonation type tends to occupy one area of the space, in a V-shaped array
- But languages do differ in exactly how they use the space for contrasts
Pitch and phonation in tone languages

• Pure phonation contrast
• Correlated pitch and phonation
• Mixed system
Relation of phonation to lexical tone in languages

- Some languages with phonation contrasts do not have lexical tone (pitch) contrasts.
- Some languages have both, cross-classifying: different tones and phonations co-occur in all possible combinations.
- Some languages use phonation as part of the tonal system: certain tones have their own correlated phonations.
- Mixed tone systems combine contrast and correlation.

Non-tonal example: Gujarati modal vs. breathy voice

<table>
<thead>
<tr>
<th>Orthography</th>
<th>Dictionary transcription</th>
<th>IPA</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>હન</td>
<td>kan</td>
<td>kan</td>
<td>ear</td>
</tr>
<tr>
<td>હનન</td>
<td>(not listed)</td>
<td></td>
<td>Krishna</td>
</tr>
<tr>
<td>બાર</td>
<td>bar</td>
<td>bar</td>
<td>twelve</td>
</tr>
<tr>
<td>બાહર</td>
<td>bəhar</td>
<td>bəar</td>
<td>outside</td>
</tr>
<tr>
<td>બાન</td>
<td>ban</td>
<td>ban</td>
<td>arrow</td>
</tr>
<tr>
<td>બાંનુ</td>
<td>bəhanū</td>
<td>bənū</td>
<td>excuse</td>
</tr>
<tr>
<td>માલિક</td>
<td>malik</td>
<td>malik</td>
<td>boss, god</td>
</tr>
<tr>
<td>મહારાજ</td>
<td>mə.ha.raj</td>
<td>məraj</td>
<td>priest, emperor</td>
</tr>
</tbody>
</table>
Cross-classifying example: Mpi (plays by rows)

<table>
<thead>
<tr>
<th>TONE (PITCH)</th>
<th>REGULAR VOICE</th>
<th>ENGLISH</th>
<th>TENSE VOICE</th>
<th>ENGLISH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low rising</td>
<td>si</td>
<td>‘to be putrid’</td>
<td>si</td>
<td>‘to be dried up’</td>
</tr>
<tr>
<td>Low level</td>
<td>si</td>
<td>‘blood’</td>
<td>si</td>
<td>‘seven’</td>
</tr>
<tr>
<td>Mid rising</td>
<td>si</td>
<td>‘to roll rope’</td>
<td>si</td>
<td>‘to smoke’</td>
</tr>
<tr>
<td>Mid level</td>
<td>si</td>
<td>(a color)</td>
<td>si</td>
<td>(classifier)</td>
</tr>
<tr>
<td>High falling</td>
<td>si</td>
<td>‘to die’</td>
<td>si</td>
<td>(name)</td>
</tr>
<tr>
<td>High level</td>
<td>si</td>
<td>‘four’</td>
<td>si</td>
<td>(name)</td>
</tr>
</tbody>
</table>
Cross-classifying example: Mazatec

<table>
<thead>
<tr>
<th></th>
<th>Creaky</th>
<th>Modal</th>
<th>Breathy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>F4</td>
<td>F6</td>
<td>F4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>F6</td>
</tr>
<tr>
<td>Mid</td>
<td>F4</td>
<td>F6</td>
<td>F6</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>F6</td>
</tr>
<tr>
<td>High</td>
<td>F4</td>
<td>F4</td>
<td>F6</td>
</tr>
</tbody>
</table>

Garellek & Keating 2011
Correlated example: Mandarin creaky voice tone

- It seems that in many languages, the lowest-pitch tone can be produced with creaky voice, or at least laryngealization.

- E.g. Mandarin Tone 3 – Kuang (2013) found that 12 speakers produced 60/60 tokens with creak (and 39/60 of Tone 4),

- See also Hockett, 1947; Chao, 1956; Davison, 1991; Belotel-Grenié & Grenié, 1994, 2004.
Mandarin example

- Female speaker
- Minimal tone set:
  - Tone 1: High 師
  - Tone 2: Rising 十
  - Tone 3: Low 使 (creaky at the end)
  - Tone 4: Falling 示 (creaky at the end)
- 3 times each

Keating & Kuo 2012
Mixed system example:
Santa Ana del Valle Zapotec

- Modal High and Rising tones
- Breathy and creaky Falling tones
- Triple with Modal-High:
  - Modal: ‘can’  lat
  - Breathy: ‘place’  -lat
  - Creaky: ‘field’  -lat-s
White Hmong tones

<table>
<thead>
<tr>
<th>Tone</th>
<th>Orthographic tone symbol</th>
<th>Example (IPA)</th>
<th>Example in Hmong orthography with English meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>High-rising (45)</td>
<td>-b</td>
<td>[pɔ́ɹ]</td>
<td>pob “ball”</td>
</tr>
<tr>
<td>Mid (33)</td>
<td>@</td>
<td>[pɔɹ]</td>
<td>po “spleen”</td>
</tr>
<tr>
<td>Low (22)</td>
<td>-s</td>
<td>[pɔɹ]</td>
<td>pos “thom”</td>
</tr>
<tr>
<td>High-falling (52)</td>
<td>-j</td>
<td>[pɔɹɎ]</td>
<td>poj “female”</td>
</tr>
<tr>
<td>Mid-rising (24)</td>
<td>-v</td>
<td>[pɔɹɭ]</td>
<td>pov “to throw”</td>
</tr>
<tr>
<td>Low-falling creaky (21)</td>
<td>-m</td>
<td>[pɔɹɭ]</td>
<td>pom “to see”</td>
</tr>
<tr>
<td>Mid-to high-falling breathy (52 or 42)</td>
<td>-g</td>
<td>[pɔɹɭ]</td>
<td>pog “grandmother”</td>
</tr>
</tbody>
</table>

Ratliff 1992, Esposito 2012
Summary, tone/phonation

- In tone languages, pitch and phonation can be independent or correlated, even within one language.

- Question: When pitch and phonation are correlated, do listeners use both kinds of information in recognizing tones?
Correlated creaky voice can help in perceiving low tones

- Creaky Tone 3 *speeds up* judgment, but doesn’t affect accuracy, which is at ceiling
- Creak helps distinguish synthesized Tone 3 from Tone 2
- Cantonese: creaky stimuli perceived more often as low tone (T4)

Belotel-Grenié & Grenié 1997; Yang 211; Yu & Lam 2011
Hmong perception experiment

- White Hmong minimal set
- Breathy-, creaky-, and modal-tone tokens had their F0 and duration modified by PSOLA re-synthesis
  - originally-breathy words now shortened and/or with lowered/falling F0
  - originally-creaky words now lengthened and/or with raised F0
  - originally-modal words now with varying duration and/or F0
- Original phonation was never modified
Stimulus examples

Natural and manipulated tokens of:

- original breathy
- original modal
- original creaky

15 White Hmong listeners identified words
Results

- They heard *breathy-tone* words only for stimuli made from an *original breathy* token; 
  F0 did not matter.

- In contrast, they heard more *creaky-tone* words when stimulus F0 was low-falling and duration was short – even if originally modal; phonation did not matter.
So, 2 different outcomes

- Breathy tone is heard when the stimulus is breathy, regardless of F0:
  phonation is criterial for the breathy tone
- “Creaky” tone is heard when the stimulus is low-pitched/short, regardless of modal/creaky phonation:
  phonation is NOT criterial for the “creaky” tone (it’s primarily a pitch contrast)
Comparison of tones

<table>
<thead>
<tr>
<th></th>
<th>Breathy</th>
<th>Modal</th>
<th>Creaky</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hmong</strong></td>
<td>pɔ̂^{42}</td>
<td>pɔ̂^{52}</td>
<td>pɔ̂^{21}</td>
</tr>
</tbody>
</table>

- **phonation contrast**
- **pitch (and duration) contrast**
Voice quality in relation to voice pitch

- Generally, phonation varies with pitch
- Speakers vary how their vocal folds vibrate, to help them vibrate faster or slower
- Speakers can thus reach higher and lower pitches than would otherwise be comfortable
Experiment on full F0 range

- Audio recordings of pitch glides up or down by English and Mandarin men and women, on vowel [a]
- On glides down, speakers told either that creak is ok, or creak is not ok
- Examples: 🔊 🔊
- Measure voice quality as pitch changes within each glide – next slide shows 2 acoustic measures
## 2 acoustic measures vs. F0

<table>
<thead>
<tr>
<th></th>
<th>Women</th>
<th>Men</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>H1-H2</strong></td>
<td><img src="#" alt="Graph" /></td>
<td><img src="#" alt="Graph" /></td>
</tr>
<tr>
<td>Breathier</td>
<td><img src="#" alt="Graph" /></td>
<td><img src="#" alt="Graph" /></td>
</tr>
<tr>
<td><strong>H1-A1</strong></td>
<td><img src="#" alt="Graph" /></td>
<td><img src="#" alt="Graph" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pitch (F0)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>300</td>
</tr>
<tr>
<td>200</td>
<td>400</td>
</tr>
</tbody>
</table>

Breathier
Red $\Delta$ = falling pitch (don’t creak)
Black $\circ$ = falling pitch (creak is ok)
Green $+$ = rising pitch  

Time runs left-to-right

Time runs right-to-left

<table>
<thead>
<tr>
<th></th>
<th>Women</th>
<th>Men</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1-H2</td>
<td><img src="image1" alt="Graph" /></td>
<td><img src="image2" alt="Graph" /></td>
</tr>
<tr>
<td>H1-A1</td>
<td><img src="image3" alt="Graph" /></td>
<td><img src="image4" alt="Graph" /></td>
</tr>
</tbody>
</table>

Breathier
Red $\triangle = \text{falling pitch (don’t creak)}$
Black $\circ = \text{falling pitch (creak is ok)}$
Green $+ = \text{rising pitch}$

<table>
<thead>
<tr>
<th></th>
<th>Women</th>
<th>Men</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>H1-H2</strong></td>
<td><img src="image1" alt="Graph" /></td>
<td><img src="image2" alt="Graph" /></td>
</tr>
<tr>
<td><strong>H1-A1</strong></td>
<td><img src="image3" alt="Graph" /></td>
<td><img src="image4" alt="Graph" /></td>
</tr>
</tbody>
</table>

Breathier

Pitch (F0)
Mandarin creaky tones: F0 at break into creak

![Bar chart showing F0 levels for T3 and T4 tones with male (M) and female (F) speakers.]
Independent tone and phonation?

- Tone and phonation contrasts can be independent, combining orthogonally within a single language.
- In these languages, speakers must largely decouple pitch and quality, so that any tone can occur with any phonation.
- How well can they do this – how phonetically independent are these phonological contrasts?
Yi languages: cross-classifying tense vs lax with tones

<table>
<thead>
<tr>
<th></th>
<th>Low tone</th>
<th>Mid tone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lax phonation</td>
<td>bə́́ ²¹  (mountain)</td>
<td>bə́́ ³³  (fight)</td>
</tr>
<tr>
<td>Tense phonation</td>
<td>bə́́ ²¹  (foot)</td>
<td>bə́́ ³³  (shoot)</td>
</tr>
</tbody>
</table>

Example from Southern Yi

Kuang 2011, Kuang & Keating 2014
Electroglottography (EGG)

speech waveform

EGG waveform

more

less

contact

Fabre 1957, Fourcin 1974; Esling 1984
Sample Yi EGG cycles: tense (top) and lax (bottom)
EGG measure: Contact Quotient (CQ)

- A measure of relative (proportional) amount of greater vs. lesser vocal fold contact

- High CQ ≈ overall more glottal constriction (higher CQ in tense or creaky voice)

Rothenberg & Mahshie 1988, Herbst & Ternstrom 2006
CQ example: White Hmong

EGG waveforms of 3 phonations

Breathy:
CQ = .41

Modal:
CQ = .57

Creaky:
CQ = .65
Independence of pitch and phonation in Yi

No tone effect on CQ

No phonation effect on F0

CQ is greater for tense (red) than for lax (blue) phonation, as expected, but tones have same CQ

F0 is greater for mid (right) than for low (left) tone, as expected, but phonations have same F0

Kuang & Keating 2012, Kuang 2013
Voice quality generally varies with voice pitch, allowing pitch-range expansion. But this is not necessary – voice quality and pitch can be quite independent in languages that cross-classify tone and phonation contrasts (e.g. Yi languages).
Conclusions

- New tools for analysis of voice quality make large-scale phonetic descriptions possible.
- The cross-language phonetic space for phonations is based on the low-frequency harmonic spectrum.
- Phonation can be correlated with pitch in tone- and non-tone languages, or independent of pitch in tone languages.
Some of my collaborators

Jody Kreiman
UCLA Head&Neck

Abeer Alwan
UCLA Engineering

Jianjing Kuang
U Penn

Marc Garellek
UCSD

Sameer Khan
Reed College

Christina Esposito
Macalester College

Yen-Liang Shue
Dolby Australia

Grace Kuo
Concordia U