Registers in tonal contrasts

Jianjing Kuang
UCLA Linguistics Department
Possible contrasting levels?

- IPA (2005) and Y-R Chao: five
- Tonal theories (e.g. Clements 1979 and Yip 1980): Two ~ three
  - four $\rightarrow$ [+/--upper] register
  - Cantonese:
    - 11 [-upper, L]
    - 22 [-upper, H]
    - 33[+upper, L]
    - 55[+upper, H]
registers

○ Tonal registers: [+/- upper] (Yip 1980, 2002)
  ● underlying distinctive pitch ranges
    e.g. Cantonese
    11 [-upper, L]  33[+upper, L]
    22 [-upper, H]  55[+upper, H]

○ Phonation-based registers
  ● E.g. Wu dialects

These two are historically related
Mental reality

However, [+/--upper] tonal registers are not perceivable cues:

- E.g. Cantonese: 22 and 33 are the most confusable tonal pair (Mok and Wong 2011)
Five-level-tone contrast is very hard to maintain, because…

Limitation in production:

- pitch range of normal speech is around 100Hz (Baken and Orlikoff 2000)
- Also see next slide, our data
Pitch range across languages (male speakers)

UCLA languages corpus
Limitation in perception:

- JND of pitch in lexical tones is about 9Hz (Silverman 2003), but a phonological contrast requires much greater difference.

- 20-30Hz difference for a tonal contrast is a small number, e.g. Cantonese 22 and 33 are very confusable and merging (Mok and Wong 2011).
Even a three-level contrast is very hard to maintain in a 100 Hz range, not to mention a fourth or fifth level.
Dispersion

○ Seems very dispreferred contrast system (Lindblom and Maddieson 1988, Flemming 2002)

○ Violate both goals: maximize perceptual contrasts; minimize articulatory efforts
This talk

Given normal hearing and speaking ability, how can native speakers produce and hear multiple contrasting level tones?
This talk

Given normal hearing and speaking ability, how can native speakers produce and hear multiple contrasting level tones?

→ The tonal production and perception of a language with five-level-tone contrasts.
Black Miao dialect, called Qingjiang Miao (Ch'ing Chiang Miao). This dialect is spoken at Shidong Kou (Shih-Tung-K'ou), Taijiang (T'ai-Kung) county of Guizhou (Kweichow) province in China.

First investigated by Fang-Kuei Li in 1940s, and reported in Kwan (1966) and Chang (1948)
## Tonal system

<table>
<thead>
<tr>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
<th>VI</th>
<th>VII</th>
<th>VIII</th>
</tr>
</thead>
<tbody>
<tr>
<td>44</td>
<td>51</td>
<td>55</td>
<td>22</td>
<td>45</td>
<td>33</td>
<td>13</td>
<td>11</td>
</tr>
</tbody>
</table>

![Graph showing tonal system with various male frequency ranges](image-url)
Tonal system

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
<th>VI</th>
<th>VII</th>
<th>VIII</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>44</td>
<td>51</td>
<td>55</td>
<td>22</td>
<td>45</td>
<td>33</td>
<td>13</td>
<td>11</td>
</tr>
</tbody>
</table>

F0 (Hz)

- Male

11
22
33
44
55

Graph showing frequency changes with different F0 values.
Stimuli:
- A minimal set of eight real monosyllabic words with [pa]. Produced by a male native speaker.
- /pa55/ "(water) full"
- /pa44/ "send"
- /pa33/ "fail"
- /pa22/ "net"
- /pa11/ "pull"
Procedure

- Familiarity phase: testing words were instructed in proper contexts
- Identification: single audio target; preceded by an audio introduction
- AX discrimination: two audio stimuli (possible pairs among eight tones); measuring RT and accuracy.
Subjects

- A total of 18 subjects, eight males and ten females, participated in this experiment. Four females, who were not native speakers of this particular Black Miao dialect, were excluded from the current analysis, leaving 14 subjects.
Hypotheses

- Tones with adjacent pitch values are in trouble
- If F0 is the only cue, accuracy for 33 should be the worst.
ID Accuracy

- T11: 80%
- T13: 90%
- T22: 10%
- T33: 80%
- T44: 60%
- T45: 100%
- T51: 90%
- T55: 70%
ID Accuracy

**Graph:**
- **Title:** ID Accuracy
- **Axes:**
  - Y-axis: 0% to 100%
  - X-axis: T11, T13, T22, T33, T44, T45, T51, T55
- **Bars:**
  - T11, T13, T33, T45, T51, T55: 80% accuracy
  - T22: 10% accuracy
  - T44: 20% accuracy
- **Legend:**
  - Red line at 80% accuracy
Dissimilarity matrix for all listeners.

<table>
<thead>
<tr>
<th></th>
<th>T11</th>
<th>T13</th>
<th>T22</th>
<th>T33</th>
<th>T44</th>
<th>T45</th>
<th>T51</th>
<th>T55</th>
</tr>
</thead>
<tbody>
<tr>
<td>T11</td>
<td>0.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T13</td>
<td>0.94</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T22</td>
<td>0.93</td>
<td>0.88</td>
<td>0.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T33</td>
<td>0.97</td>
<td>0.78</td>
<td>0.95</td>
<td>0.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T44</td>
<td>0.98</td>
<td>1.00</td>
<td>0.70</td>
<td>0.98</td>
<td>0.03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T45</td>
<td>0.94</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T51</td>
<td>0.94</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>T55</td>
<td>0.95</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.92</td>
<td>0.88</td>
<td>0.88</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Followed Johnson (2003)
Production experiment

- A wordlist of minimal monosyllabic sets for the eight tones was created based on Li's transcriptions (Kwan1966, Chang 1948). 23 minimal or near-minimal sets were confirmed by the speakers.
- Simultaneous EGG and audio recordings were then collected from 15 native speakers (ten males and five females).
Measures

- **Pitch related:**
  mean F0, F0 change, offset, onset
- **Duration**
- **Voice quality related** (* corrected):
  H1*, H2*, H4*, A1*, A2*, A3*;
  H1*-H2*/A1*/A2*/A3*;H2*-H4*
- **EGG**
  CQ (contact quotient), SQ (skew quotient), PIC, PDC
MDS Production space

Pitch measures only
- Mean F0,
- F0 change
- onset
- offset
- duration
R = 0.17

Perceptual space

Production space
MDS Production space

Pitch + voice quality
H1*, H2*, H4*
H1*-H2*
H1*-A1*/A2*/A3*
H2*-H4*
R=0.76

Perceptual space

Production space
Voice qualities in tones

11, 33 and 55 benefit from phonation cues


- /pa55/ "(water) full"
- /pa44/ "send"
- /pa33/ "fail"
- /pa22/ "net"
- /pa11/ "pull"
Discussions

- 55 and 11 can benefit from both pitch cues and phonation cues.
- For the mid-range tones that have very similar pitch cues, 33 is distinctive from 22 and 44 primarily by the phonation cue.
- 22 vs. 44, the tonal contrast with only a pitch difference is the most confusable.
Perceptual space  
Production space
Tonal registers

- Tense
- Breathy
- Modal
- Vocal fry

Pitch levels:
- Tense: 55
- Breathy: 44, 33, 22
- Modal: 44, 33, 22
- Vocal fry: 11
Contributions of non-modal phonations

- 55 and 11: enhance the tonal contrasts
- 33: distinctive cue from the other mid tones
Two types of non-modal phonations

- **Pitch-driven type**: e.g. vocal fry, falsetto and tense

- **Pitch-range production is related to certain phonation types**: (Hollien 1974, Titze 1988, Baken&Orlikoff 2000)
  - The lowest pitch range (i.e. < 70Hz) – vocal fry
  - The highest pitch range (i.e. > 175Hz for males, 275Hz for females) – falsetto/tense
Two types of non-modal phonations

- **Pitch independent type**: create an independent dimension for tonal contrasts
  - 33 vs. 22 and 44 (similar pitch but different registers are not confusible)
  - Also falling tones in Green Mong (Andruski 2006) and White Hmong (Garelleck et al. 2012)

- Breathiness contrast – relative on a phonation continuum

Phonation type

- Most open
- Voiceless
- Breathy
- modal
- Creaky
- Most closed
- Glottal closure

Ladefoged’s model of continuous breathiness
Concluding remarks

- Dispersion of the five level tones is fine:
  - Pitch-driven phonations help to produce extreme F0 targets, and thus enhance the perceptual differences for the highest and lowest tones
  - Pitch-independent phonations create an independent dimension for tonal contrasts so that tones with similar pitches are very well distinguished
  - Pure pitch contrasts are hard even for two levels
associations

- Pitch register – pitch-driven non-modal phonations – allophonic
  - Realization of extreme pitch targets (highest, lowest)
  - E.g. vocal fry: Mandarin, Cantonese falsetto: Gaoba Dong, PPhN Thai
- Phonation-based register – pitch independent non-modal phonations – phonemic
  - E.g. Yi, Mazatec
Thank you! 😊

put vot
Acknowledgments

- Advising committee: Patricia Keating, Jody Kreiman, Sun-Ah Jun, Bruce Hayes and Megha Sundra for invaluable comments
- Colleagues in UCLA phonetics lab for all the discussions
- Professor Jiangping Kong of Peking University for the equipment support
- Professor Defu Shi of Minzu University of China for introduction of Black Miao letters
- Lindy Mark for the help of getting Prof Li’s word list
- Black Miao friends who kindly participated in this study
Revisit tonal registers

- Three kinds of registers:
  - Pitch register: the frequency scale with normal height versus super high (with falsetto) or super low (with vocal fry).
  - Phonation-based register: phonation contrast in languages
  - [+/-upper] register: underlying distinctive pitch ranges e.g. Cantonese. /33/ [-low, -upper]; /22/ [+low, +upper]
  - Phonation-based and [+/-upper] are historically related

[+/-upper] registers have no mental reality.
Identification matrix for all listeners.

<table>
<thead>
<tr>
<th>Target</th>
<th>T11</th>
<th>T13</th>
<th>T22</th>
<th>T33</th>
<th>T44</th>
<th>T45</th>
<th>T51</th>
<th>T55</th>
</tr>
</thead>
<tbody>
<tr>
<td>T11</td>
<td>85%</td>
<td>3%</td>
<td>0%</td>
<td>2%</td>
<td>0%</td>
<td>2%</td>
<td>7%</td>
<td>2%</td>
</tr>
<tr>
<td>T13</td>
<td>0%</td>
<td>88%</td>
<td>0%</td>
<td>7%</td>
<td>0%</td>
<td>2%</td>
<td>3%</td>
<td>0%</td>
</tr>
<tr>
<td>T22</td>
<td>0%</td>
<td>7%</td>
<td>19%</td>
<td>5%</td>
<td>64%</td>
<td>0%</td>
<td>3%</td>
<td>2%</td>
</tr>
<tr>
<td>T33</td>
<td>0%</td>
<td>10%</td>
<td>3%</td>
<td>81%</td>
<td>0%</td>
<td>2%</td>
<td>2%</td>
<td>2%</td>
</tr>
<tr>
<td>T44</td>
<td>0%</td>
<td>2%</td>
<td>7%</td>
<td>3%</td>
<td>76%</td>
<td>0%</td>
<td>0%</td>
<td>12%</td>
</tr>
<tr>
<td>T45</td>
<td>0%</td>
<td>5%</td>
<td>2%</td>
<td>0%</td>
<td>0%</td>
<td>93%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>T51</td>
<td>3%</td>
<td>2%</td>
<td>0%</td>
<td>5%</td>
<td>0%</td>
<td>0%</td>
<td>90%</td>
<td>0%</td>
</tr>
<tr>
<td>T55</td>
<td>2%</td>
<td>0%</td>
<td>2%</td>
<td>3%</td>
<td>7%</td>
<td>0%</td>
<td>2%</td>
<td>84%</td>
</tr>
</tbody>
</table>
Preliminary results

- 5 repetitions of T3 and T4 from 6 females were coded. All the tokens presented vocal fry

  ➔ The presence of vocal fry in Mandarin is not sensitive to tonal category

- Pitch range of T3 and T4 (Next slide): no significant difference between T3 and T4 ($t(28) = -1.67, p>0.1$), both around 175-180 Hz

  ➔ The presence of vocal fry is only related to F0 values
Preliminary results – pitch values for vocal fry (6 females)
Pitch range ~ phonation

- Pitch-range production is related to certain phonation types: (Hollien 1974, Titze 1988, Baken&Orlikoff 2000)
- The lowest pitch range (i.e. < 70Hz) – vocal fry
- The highest pitch range (i.e. > 175Hz for males, 275Hz for females) – falsetto

Phonemic phonation contrasts seem best to occur with the mid pitch-range.
- Phonation contrast in high tones are limited (e.g. Yi and Mazatec)
- Phonation contrast is neutralized in the focus position (e.g. Zapotec)
Define non-modal phonations

- **Vocal fry**: adductive tension and compression of the vocal folds; strong damping occur with the pulses; naturally produced in the lowest pitch. usually also refers to period-doubling (cycles alternate in a repeating long-short-long-short pattern)

- **Falsetto**: long narrow leakage between the vocal folds, so Open Quotient is high; sinusoid-like glottal pulses; naturally produced in the highest pitch range; acoustic indicators are unclear

- **Tense**: The vocal folds get a lot of longitudinal tension; naturally happens when pitch is approaching the highest limit or when people lift heavy objects

- **Breathy voice and Creaky voice**: ambiguous linguistic categories; relative on glottal closure continuum
Cantonese

Figure 1. *F0* traces of the six lexical tones.
Pitch location judgment

The diagram shows box plots for different pitch locations, including Fhigh, Flow, Fmidhigh, Fmidlow, Mhigh, Mlow, Mmidhigh, and Mmidlow. The y-axis represents the pitch location values ranging from 200 to 800.
Contact Quotient measures

Peak Increase in Contact
Peak Decrease in Contact

Slow Quotient (Speed Quotient)

Contact Quotient (CQ)
4 methods

Peak velocities of contact
Increase and Decrease from dEGG

closing duration / opening duration from EGG
MDS plot for seven languages phonations

Dimension 1

Dimension 2

Dimension 3

BL
BT
GB
GM
HB
HC
HM
LL
LT
MaC
MaM
YiL
YiT
ZB
ZC
ZM