Laryngealization and features for Chinese tonal recognition

Kristine M. Yu
University of California, Los Angeles

The UCLA Phonetics Lab

Abstract

- Automatic tonal recognition abstracts away from **laryngealization**
- But laryngealization prevalent in **Beijing Mandarin** and **Hong Kong Cantonese** production data
  - Causes significant disturbance of f0 detection
  - Predominantly occurs in lowest tones in tonal inventory
- Moreover, human tonal perception aided by laryngealization
- **Laryngealization should be treated in acoustic front end in automatic tonal recognition**

Definition of laryngealization

- Working definition: **irregular phonation** resulting in human percept of **creaky voice**
  - Irregular phonation: "...anomaly with respect to the usual quasi-periodic behavior of the vocal folds" (Surana & Slifka 2006)
- Two main mechanisms (Gerratt & Kreiman 2001)
  - **Vocal fry**: train of low frequency glottal pulses, heavy damping between pulses
  - **Period doubling**: waveform cycles alternating in amplitude and/or frequency, bitonal percept
- **f0 not well-defined for either mechanism**

Vocal fry (Mand. T3)
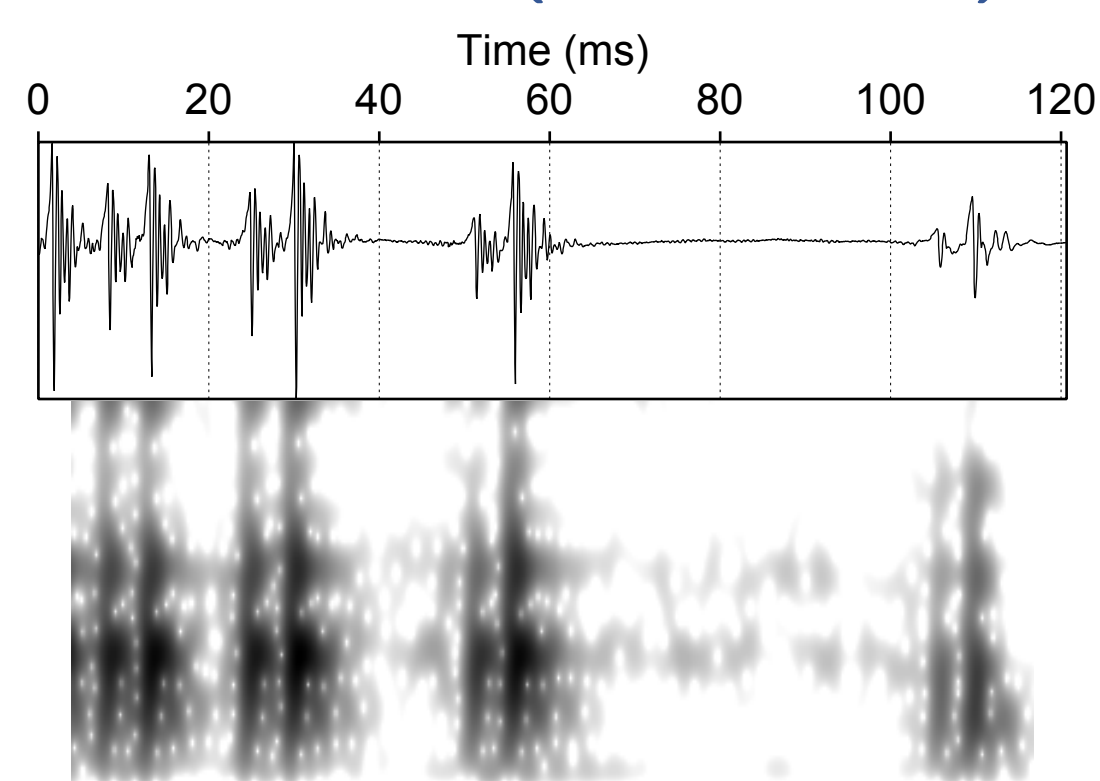
Period doubling (Cant. T4)
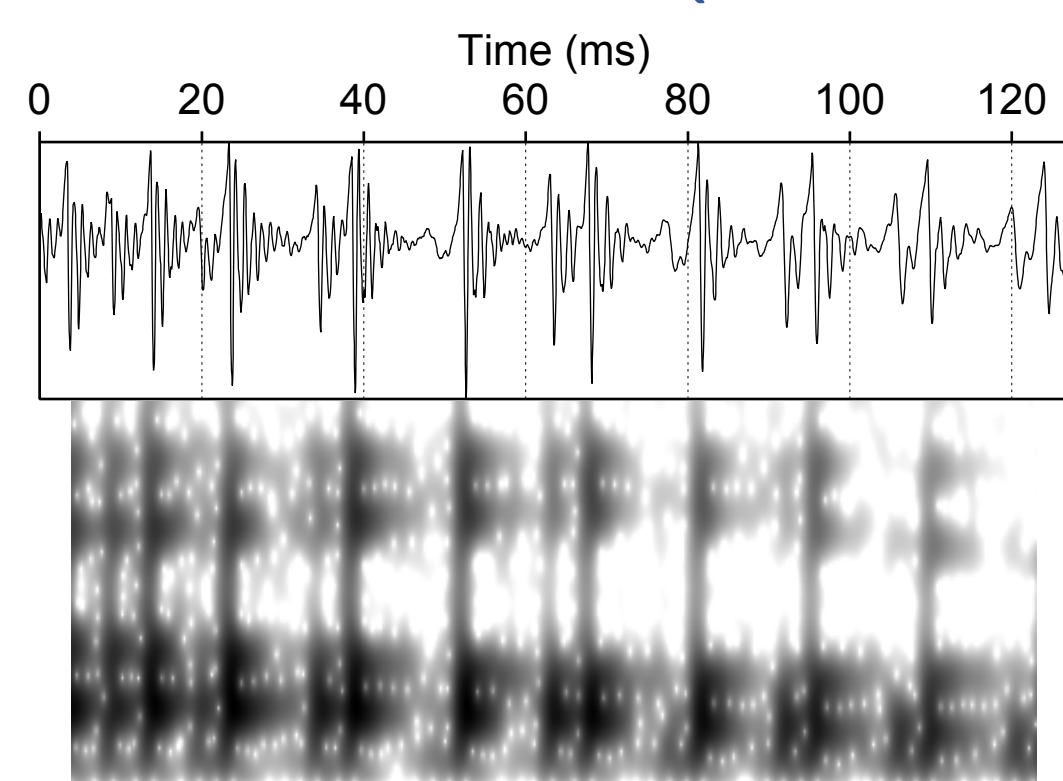
Problem: laryngealization disturbs f0 detection

- **Any tonal recognizer relies primarily on f0-based features and depends on reliable f0 tracking**
- f0 detection fails in vocal fry, period doubling

Laryngealization in Mandarin and Cantonese

- Mandarin and Cantonese: tonal languages without contrastive phonation
- **Mandarin Tone 3** (low dipping) well-known to be typically laryngealized (Davison 1991); Tone 2 (rise) and 4 (fall) can be too (Belotel-Grenié & Grenié 1997)
- **Cantonese Tone 4** (low fall) anecdotally suggested to sometimes be laryngealized (Vance 1977)

Description of tonal production corpus

- 8 speakers (4M/4F) of Beijing Mandarin, and of Hong Kong Cantonese
- **Sentence-medial bitones** of all possible tonal combinations, 5 repetitions each sentence, speech rate ≈ 3-4 syllables/second
- Bitones flanked by L/H tone targets (Mand.), mid tone (Cant.)
- Mandarin neutral tone, Cantonese checked tones excluded
- Total sentences per speaker: Cantonese 540; Mandarin 320

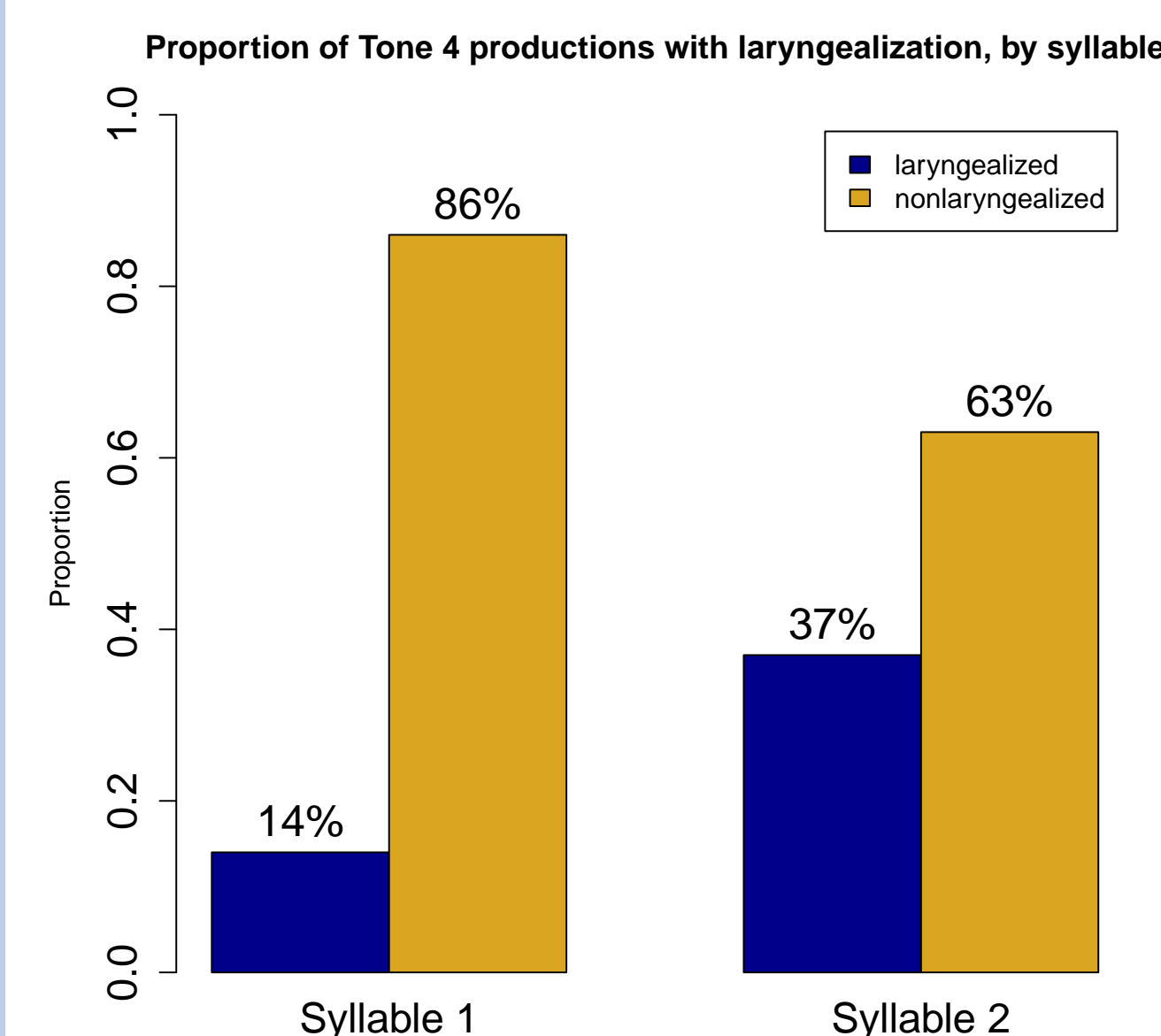## Laryngealization prevalent in production corpus

Figure: Cantonese T4 laryngealized more frequently in second syllable of bitone

- **Cantonese T4 laryngealized about 25% of time**
  - Other tones rarely laryngealized (2% for T2 (rise), less often for others)
  - Laryngealization more frequent in 2nd syllable of bitone, perhaps due to prosodic boundary
- **Mandarin T3 laryngealized about 68% of time**
  - Other tones rarely laryngealized (8% for T2 (rise), less often for others)
  - Prevalence results roughly similar to Belotel-Grenié & Grenié (1997)
- Significant individual variation
  - Some almost always laryngealize Cantonese T3/Mandarin T4
  - Some almost never laryngealize

Strategies: laryngealization in the acoustic front end

1. **Abstract away from laryngealization**
   - Average out/smooth over f0 disturbances
   - Output no f0 information (missing values)
2. **Subsume laryngealization under f0**
   - If laryngealized region ⇒ assign low f0 value
3. **Treat laryngealization independently from f0**
   - Train on laryngealization-related feature(s) in addition to f0-based features

Current automatic tonal recognizers abstract away

- Mandarin "tone nucleus" recognizer: **manual correction** of f0 contours for pitch halving or doubling (Zhang & Hirose 2004)
- Cantonese "supratone" recognizer: f0 **averaged** for each of three subsegments per syllable final (rime) (Qian, Lee, & Soong 2007)
- Surendran and Levow (2008) did investigate use of voice quality features (band energy, spectral features) in Mandarin
  - Only band energy features improved recognition (of neutral tone)
  - But spectral features based on harmonics, thus dependent on f0 detection

But human tonal recognizers use laryngealization

- Mandarin listeners need to hear less of tone before accurately identifying as T3 if creaky instance of T3 (Belotel-Grenié & Grenié 1997)
- Cantonese listeners have **higher tonal identification accuracy and faster reaction times on T4 with creaky voice** than without creaky voice: 85% vs. 62% correct (Lam and Yu 2010)
  - Monosyllables extracted from connected speech (same corpus as described here)
  - Less likely to identify creaky T4s (low fall) as T6 (low level)

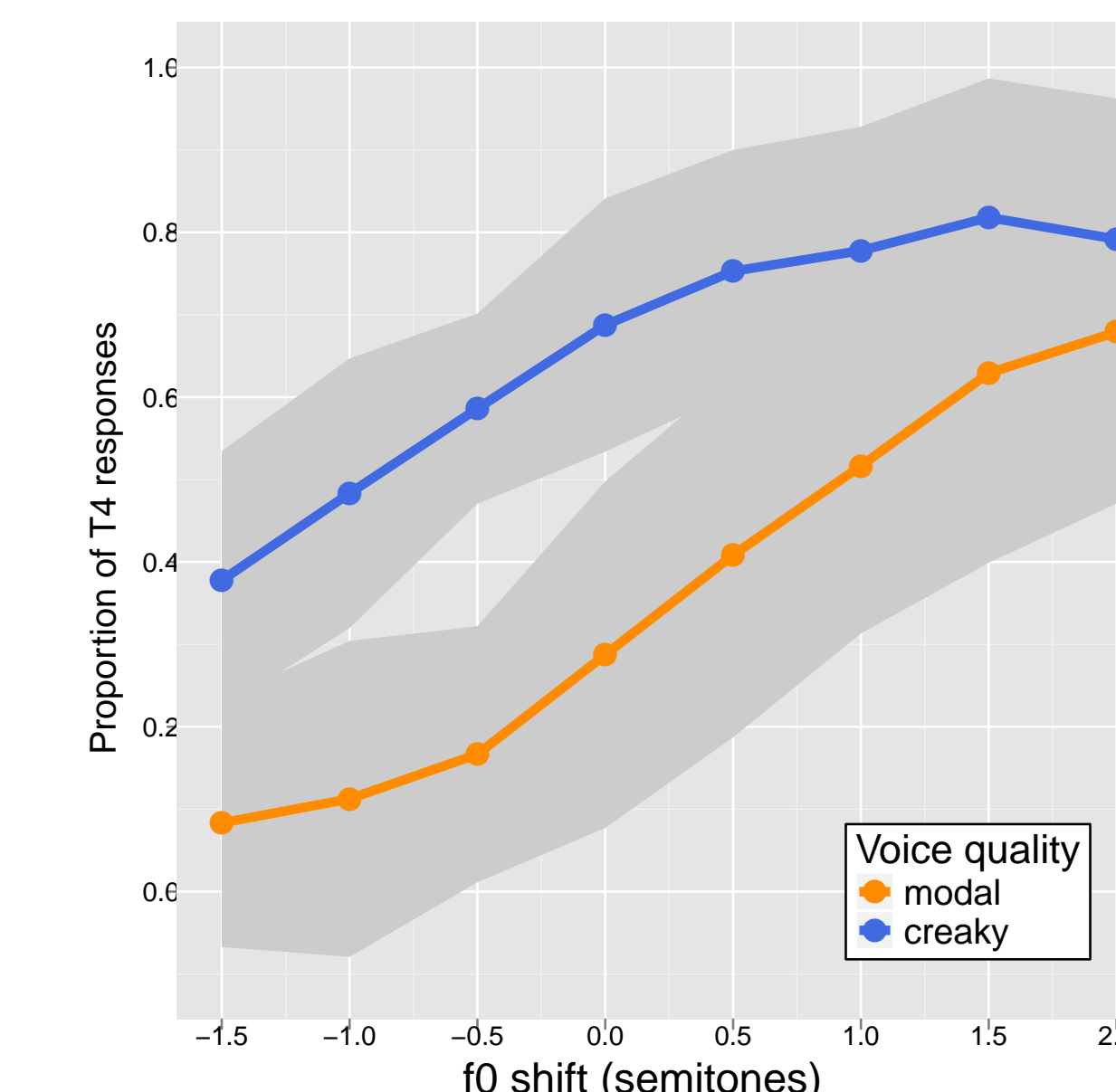## ...and they don't subsume laryngealization under f0

Figure: Cantonese listeners biased toward Tone 4 perception by laryngealization

- 20 Cantonese listeners asked to identify syllable as T4 or T6, given preceding syllable
- f0 of preceding syllable shifted in 8 step continuum to provide differing relative f0 contexts (Wong & Diehl 2003)
- Target syllable cross-spliced with period-doubled phonation
- **Period doubling biases listeners towards T4 percept, but does not globally override f0 information**

Conclusions

- High frequency of laryngealization makes abstracting away from laryngealization in automatic tonal recognition difficult
  - Laryngealization sometimes over entire syllable ⇒ not reasonable to average out/smooth over f0 tracking irregularities
- In fact evidence from human speech production and perception suggests laryngealization can aid tonal recognition
  - Distribution heavily conditioned on tone category—mostly in low tones
  - Aids and biases human tonal recognition in addition to f0 cues
- **Research in automatic tonal recognition should reconsider use of voice quality features**

Acknowledgments