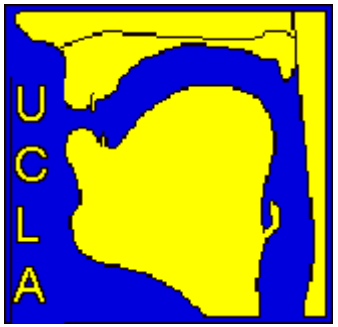


Some factors affecting voice quality within and between speakers

Pat Keating

UCLA Linguistics Department

Phonation

back

front

- **Phonation**: sound production in the larynx, usually by vocal fold vibration (**voice**, or **voicing**)
- How fast the folds vibrate (the F0) 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 

Cassette with Laver 1980, *The Phonetic Description of Voice Quality*

This talk: some factors that give rise to voice variation

- Variation in voice pitch
- Related prosodic variation
- Coarticulation from consonants
- Difficulties with voiced consonants
- Differences across individuals

Some of my collaborators

Jody Kreiman
UCLA Head&Neck

Marc Garellek
UCSD

Jianjing Kuang
U Penn

Grace Kuo
Concordia U

Soo Jin Park
UCLA Engineering

Megan Risdal
Formerly UCLA

Yen-Liang Shue
Dolby Australia

Caroline Sigouin
U Laval

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

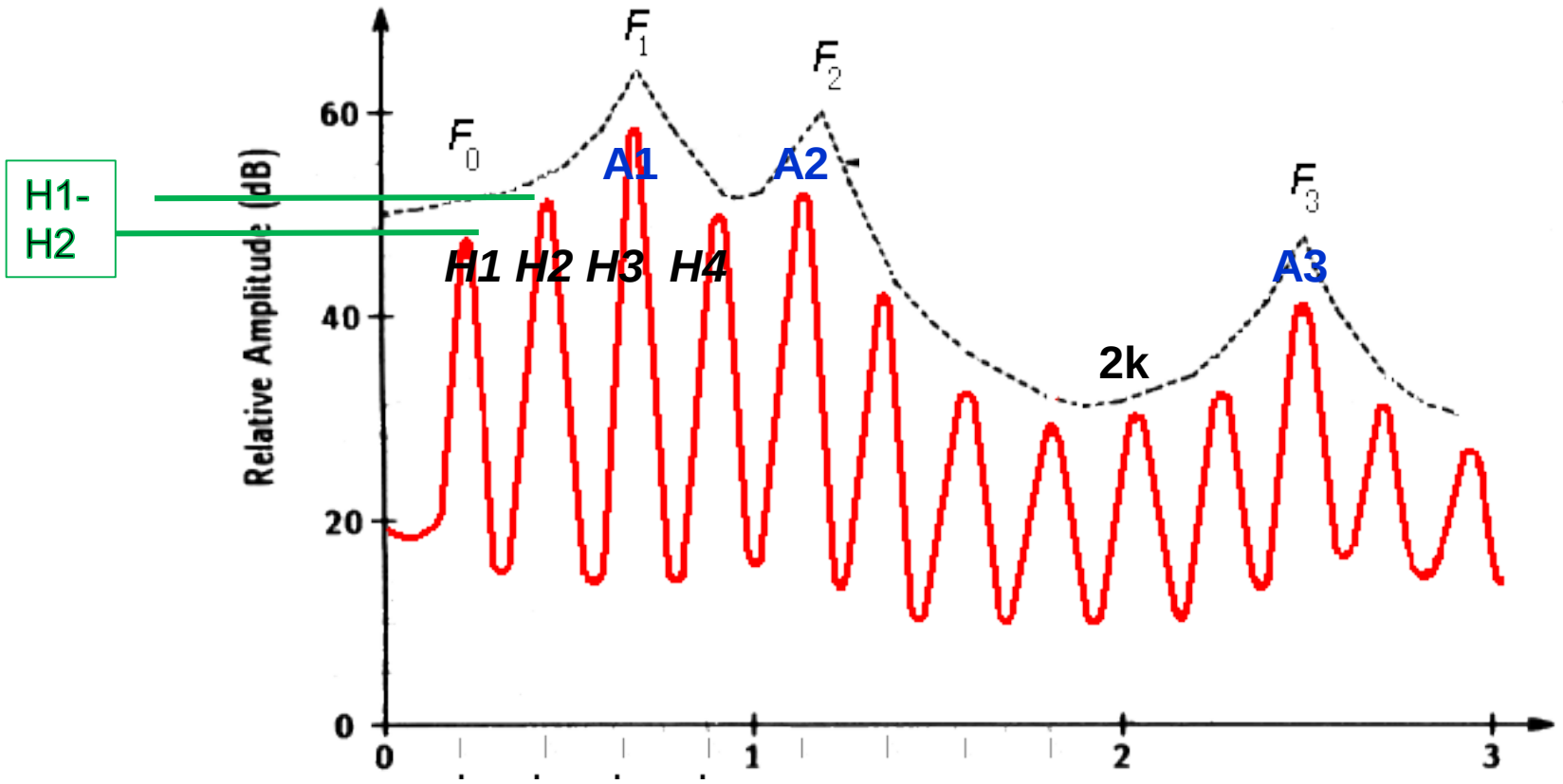
VoiceSauce measures

- F0 from STRAIGHT, Snack, or Praat
- H1, H2, H4
- 2000 Hz, 5000 Hz
- 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-H2k(*)
- H2k(*)-H5k
- Energy
- Subharmonic to Harm. Ratio
- Cepstral Peak Prominence
- Harmonic to Noise Ratios (4 freq. bands)
- Strength of Excitation

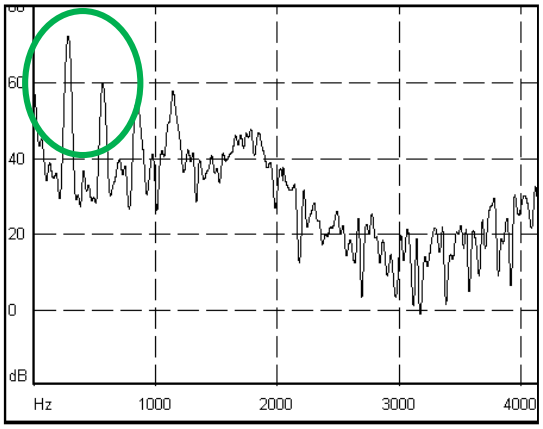
Acoustic measures based on harmonics in spectrum



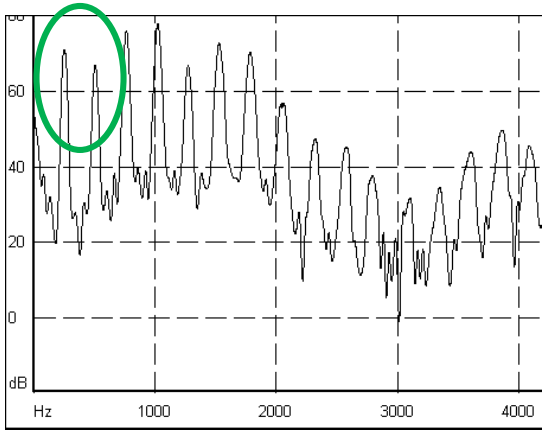
Shue 2010, Kreiman et al. 2007, Kreiman et al 2014

H1-H2 example: Jalapa Mazatec

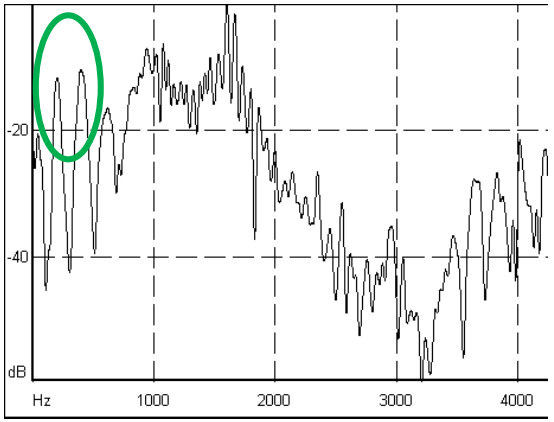
breathy



modal



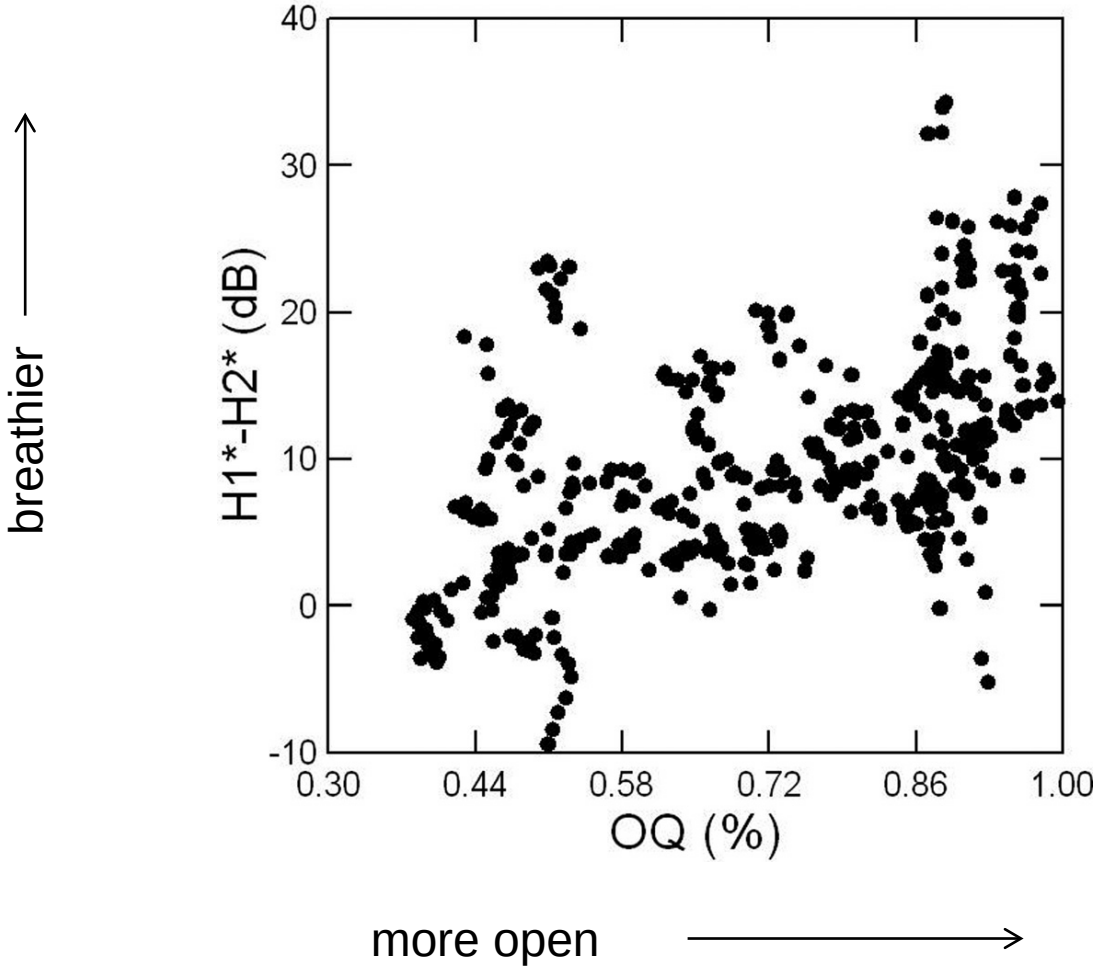
creaky



Breathy	Modal	Creaky
b̤a^{34}	ba^{32}	b̰a^{3}

Kirk et al. 1984, Garellek & Keating 2011

H1-H2 is partially related to how long the glottis is open during each cycle of voicing



Kreiman et al. 2012, Chen et al. 2013

Ladefoged's continuum: size of glottal opening

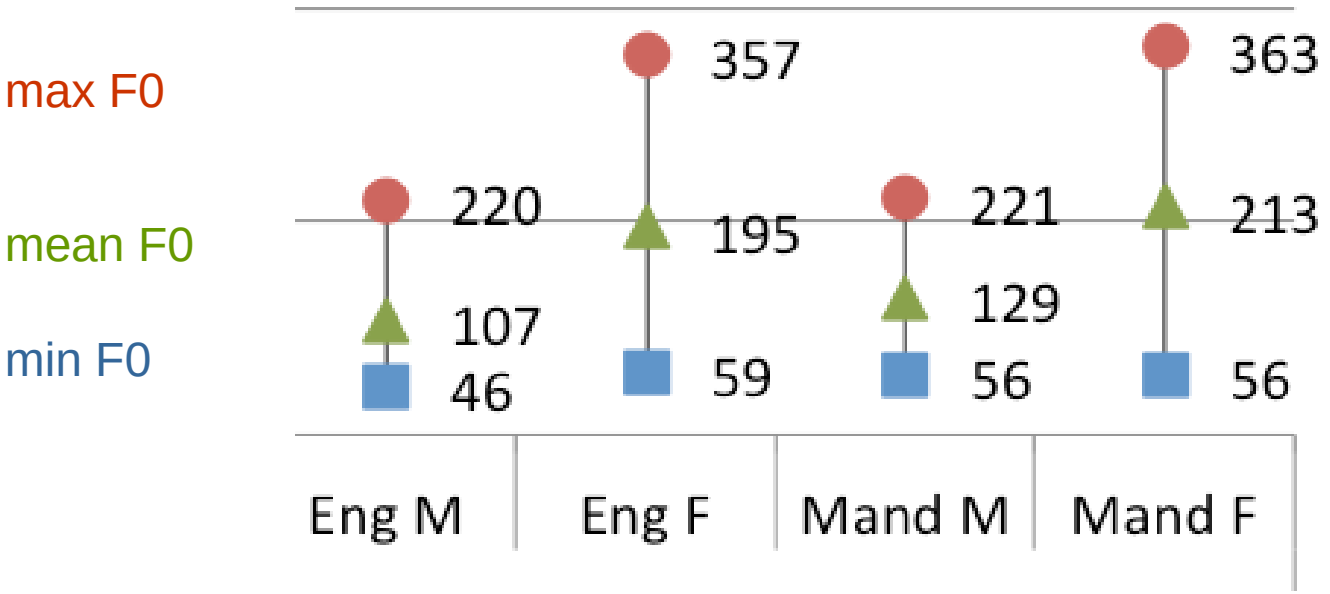
	Most open ◄			► Most closed
Phonation type	Voiceless	Breathy	Modal	Creaky	Glottal closure

Ladefoged 1971, Gordon & Ladefoged 2001

Sources of voice variation

- Variation in voice pitch (F0)
- Related prosodic variation
- Coarticulation from consonants
- Difficulties with voiced consonants
- Differences across individuals

Voice pitch variation: mean F0 ranges in “Rainbow Passage”



English- and Mandarin-speaking men and women

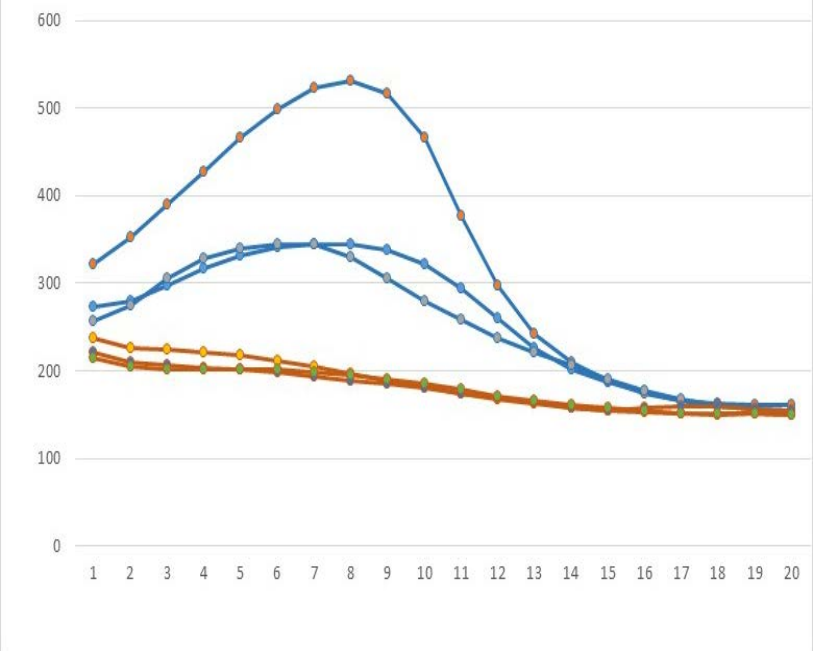
Keating & Kuo 2012

Voice quality in relation to voice pitch

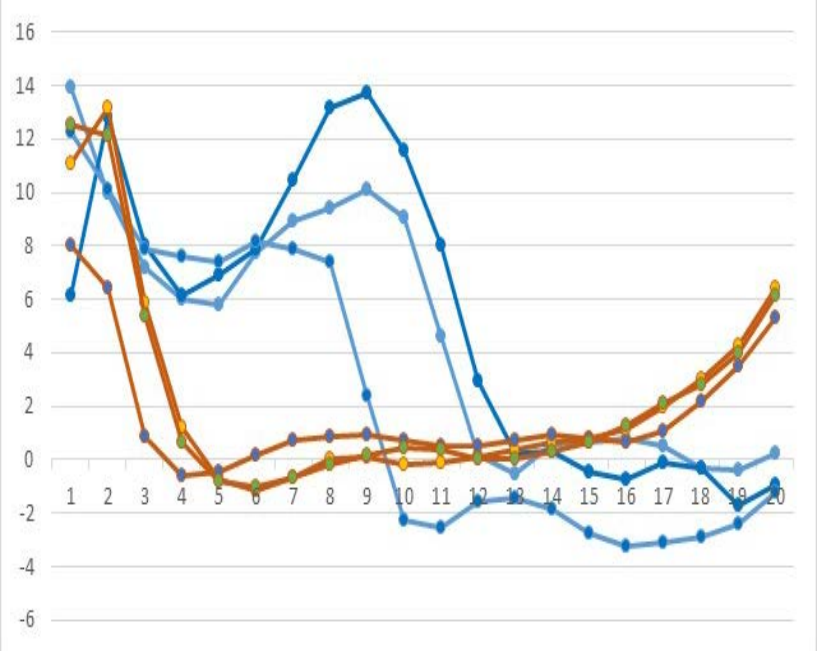
- Generally, phonation varies with pitch
- Speakers vary *how* their vocal folds vibrate, to help them vibrate faster or slower
- Speakers can thus reach higher and lower pitches than would otherwise be comfortable

Example: 3 tokens of “sure” at lower pitch and higher

Voice pitch (F0)



Voice quality (H1*-H2*)

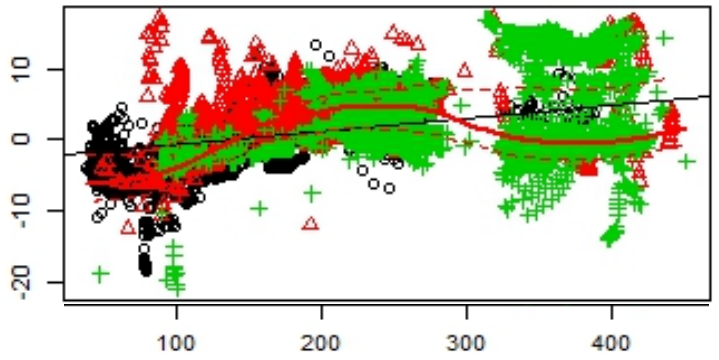
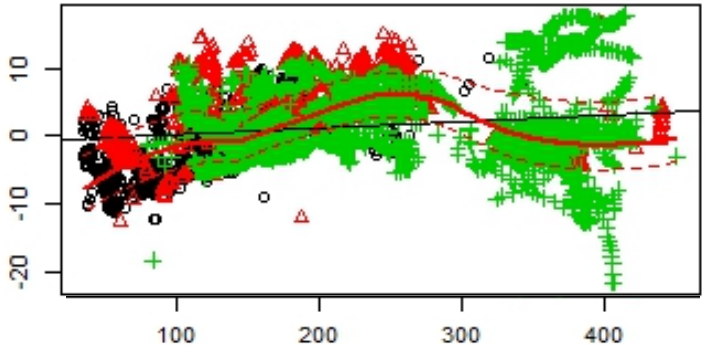
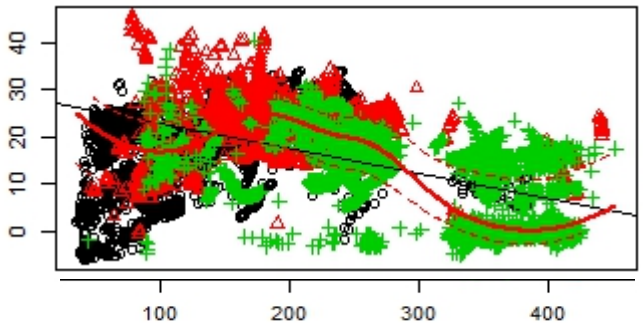
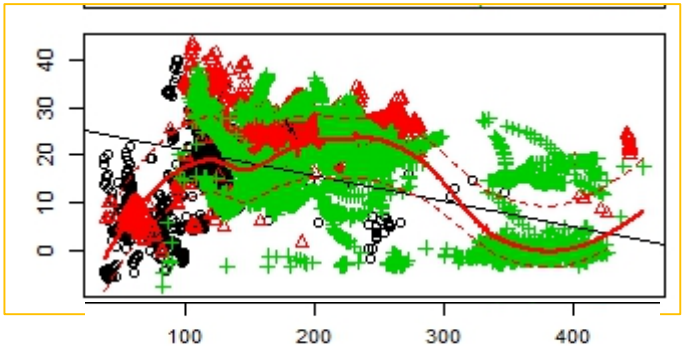


Keating & Kuo 2012

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

	Women	Men
H1*-H2*		
H1*-A1		

Breathier

Pitch (F0)

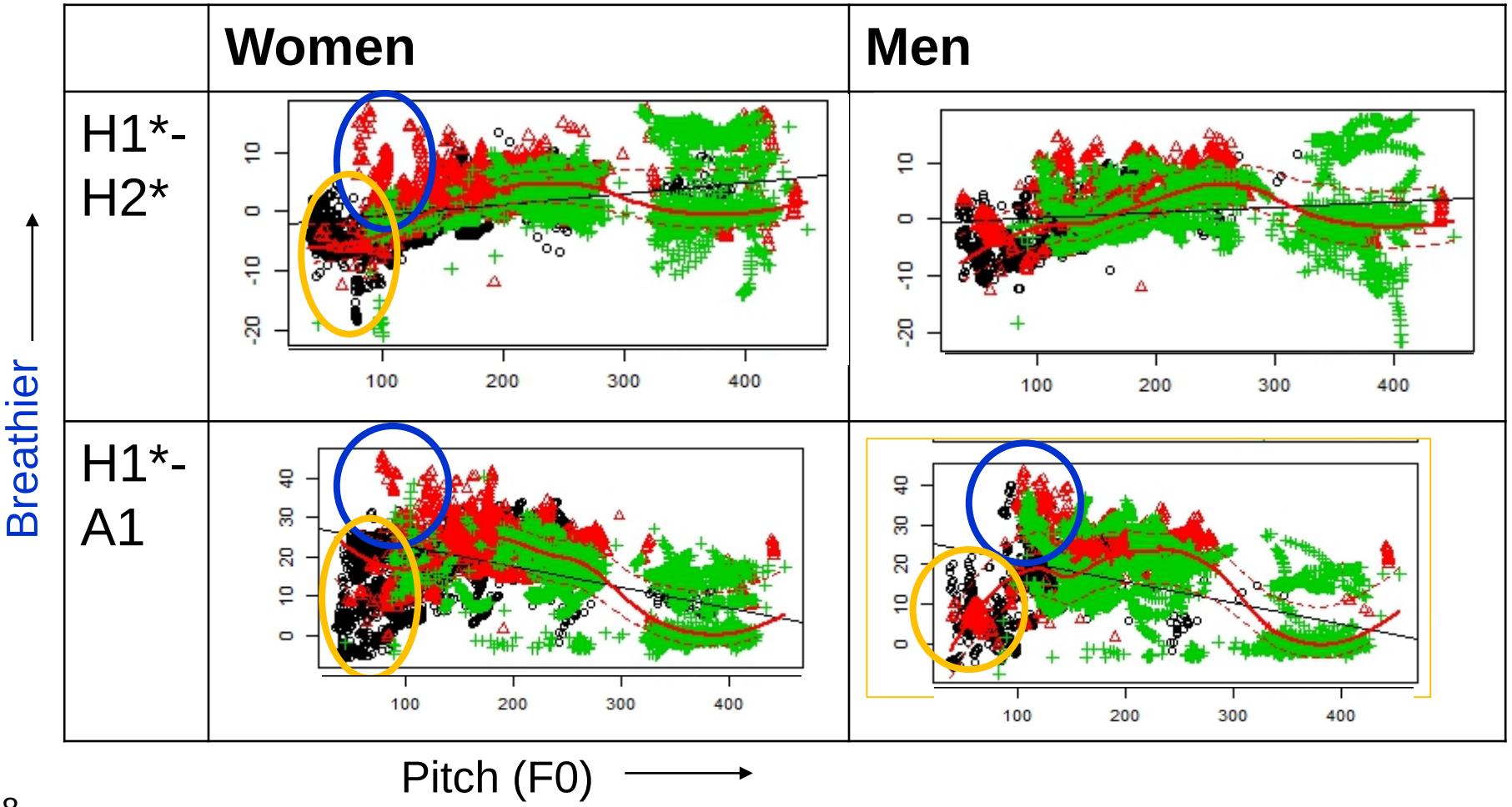
Red Δ = falling pitch (don't creak)

Black ○ = falling pitch (creak is ok)

} Time runs right-to-left

Green + = rising pitch

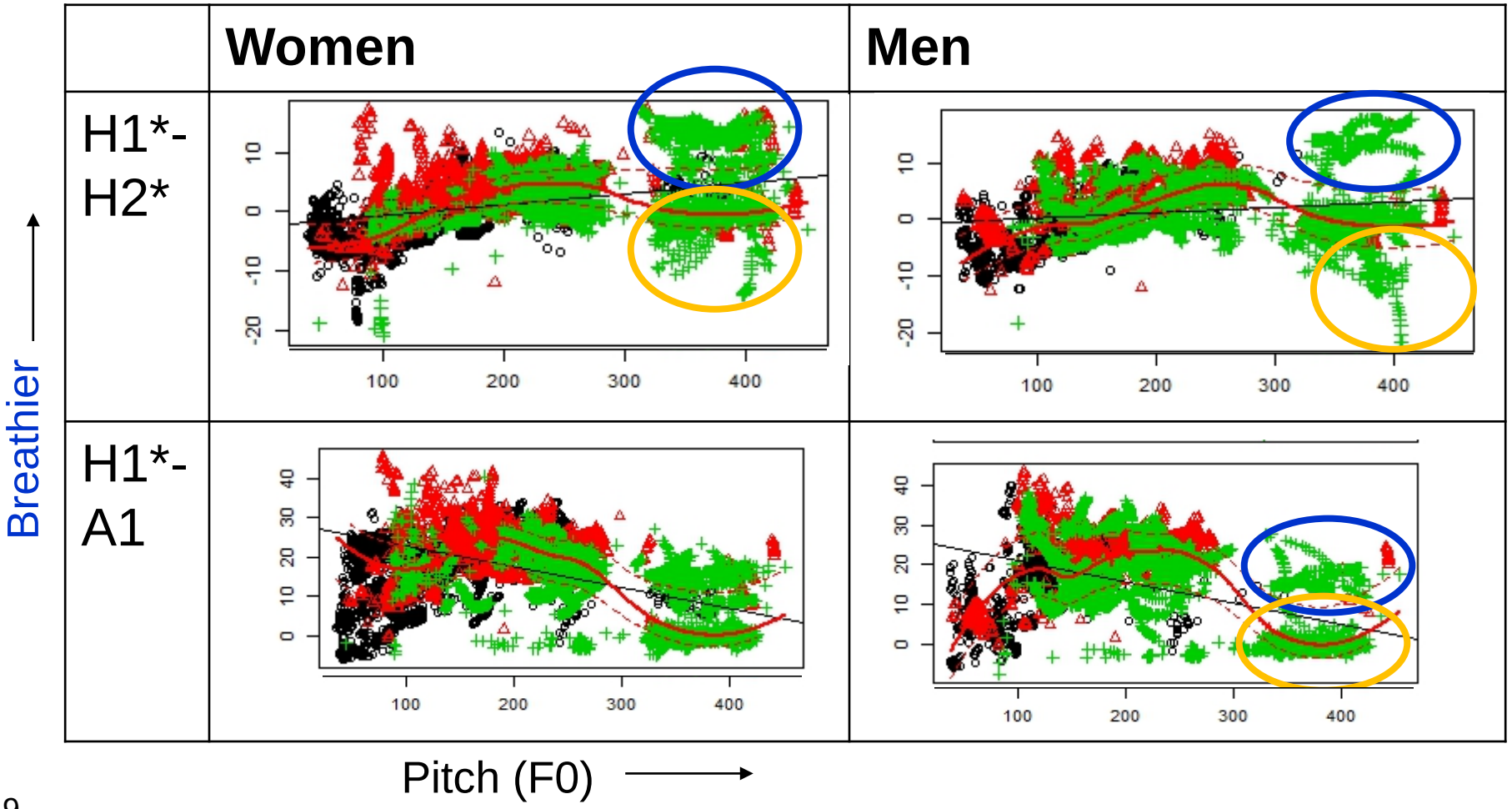
} Time runs left-to-right



Red Δ = falling pitch (don't creak)

Black ○ = falling pitch (creak is ok)

Green + = rising pitch



Falling into creak

Thanks to Jody Kreiman, Bruce Gerratt, and Henry Tehrani for help making high-speed videos

Sources of voice variation

- Variation in voice pitch
- Related prosodic variation
- Coarticulation from consonants
- Difficulties with voiced consonants
- Differences across individuals

Mandarin: creaky voice, tones, pitch

- 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), creaking at the same F0 in the 2 tones
- See also Hockett, 1947; Chao, 1956; Davison, 1991; Belotel-Grenié & Grenié, 1994, 2004

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

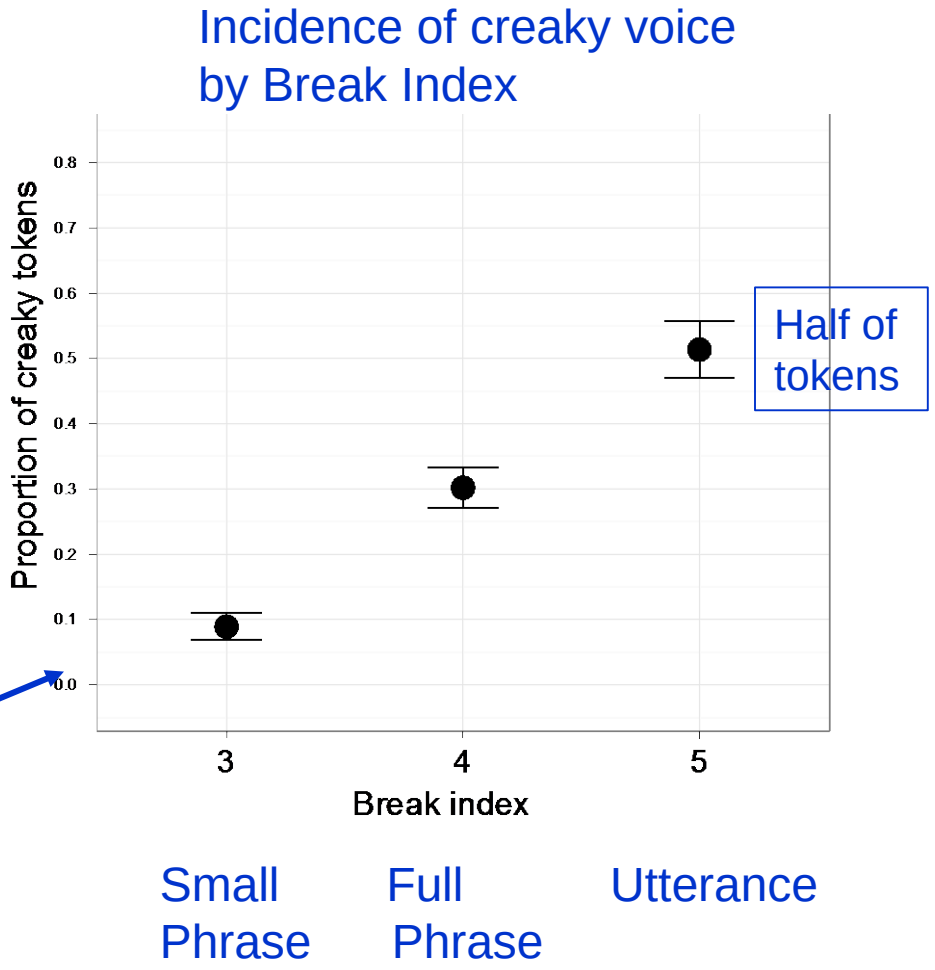
Creaky voice can help in perceiving low tones

- 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

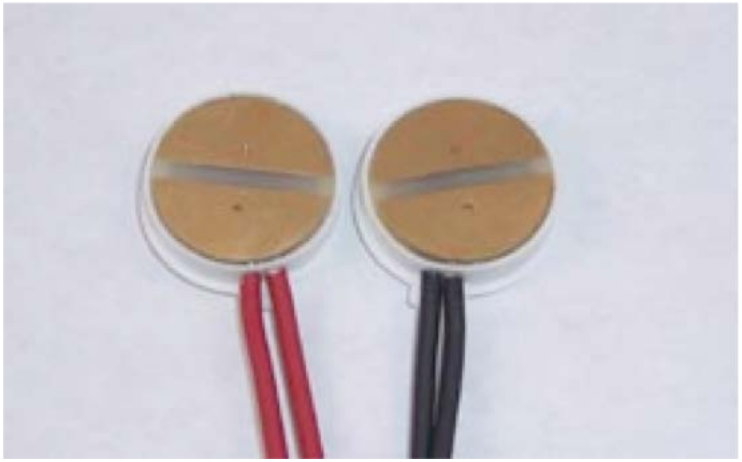
Relation to “phrase-final creak” in English

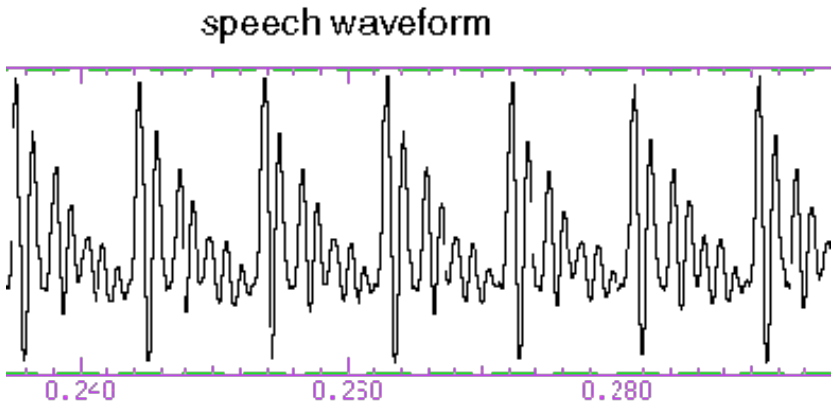
- final creak in the BU Radio Corpus, before different kinds of phrase breaks
- only 2 factors favor creak there:
- the lower the pitch and the bigger the phrase break, the more likely is creak

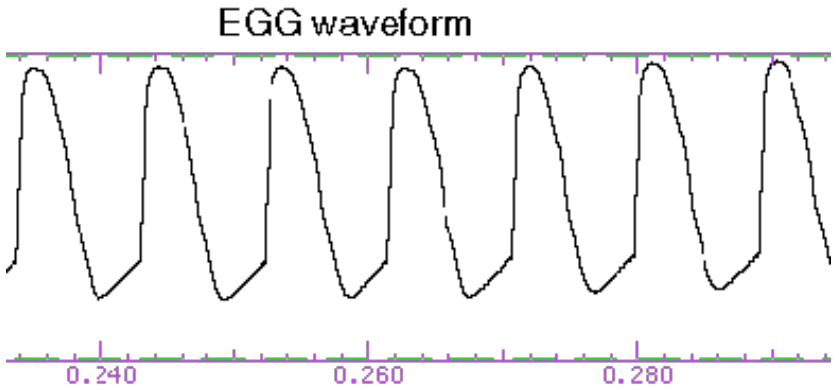


Garellek & Keating 2015

Electroglottography (EGG)







more

less

contact

Fabre 1957, Fourcin 1974; Esling 1984

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)

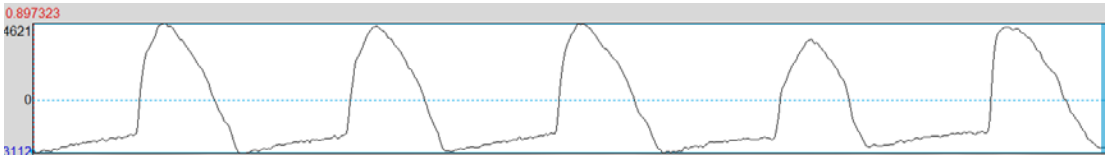
Rothenberg & Mahshie 1988, Herbst & Ternstrom 2006

CQ example: White Hmong

EGG waveforms of 3 phonations

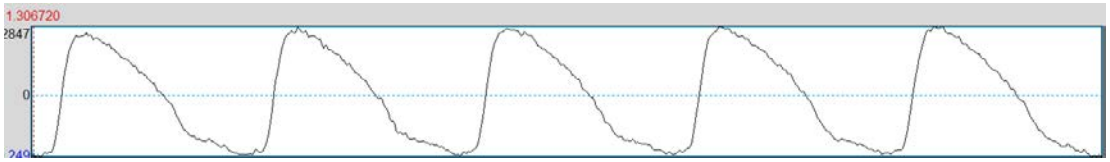
Breathy:

CQ = .41



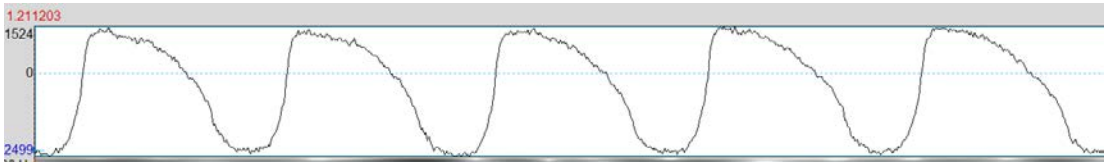
Modal:

CQ = .57



Creaky:

CQ = .65

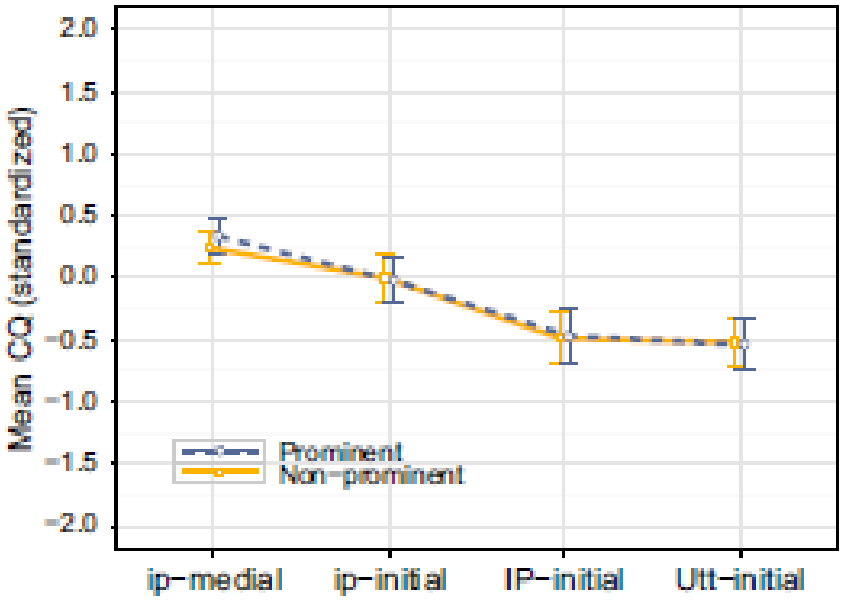


more contact

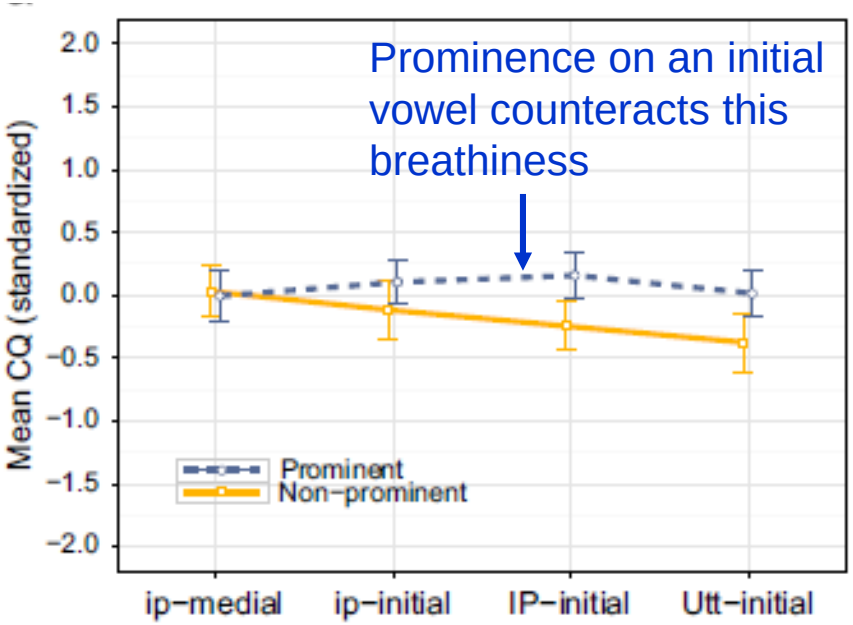
less contact

English phrase-initial phonation: breathier (lower CQ) in higher prosodic domains

Initial sonorants



Initial vowels



higher prosodic domain ⟶

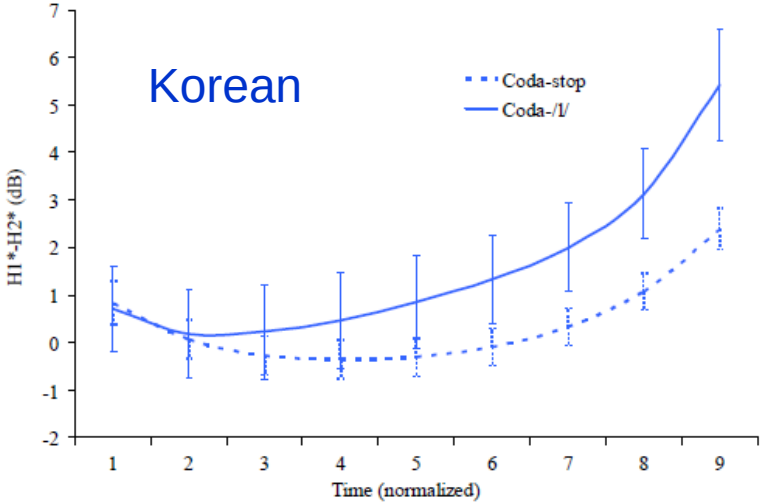
Garellek 2014

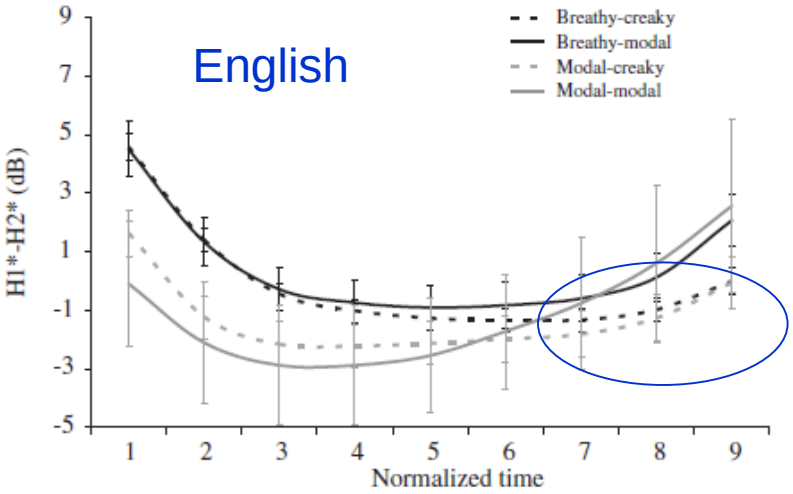
Sources of voice variation

- Variation in voice pitch
- Related prosodic variation
- Coarticulation from consonants
- Difficulties with voiced consonants
- Differences across individuals

Coda glottalization makes previous vowel creaky

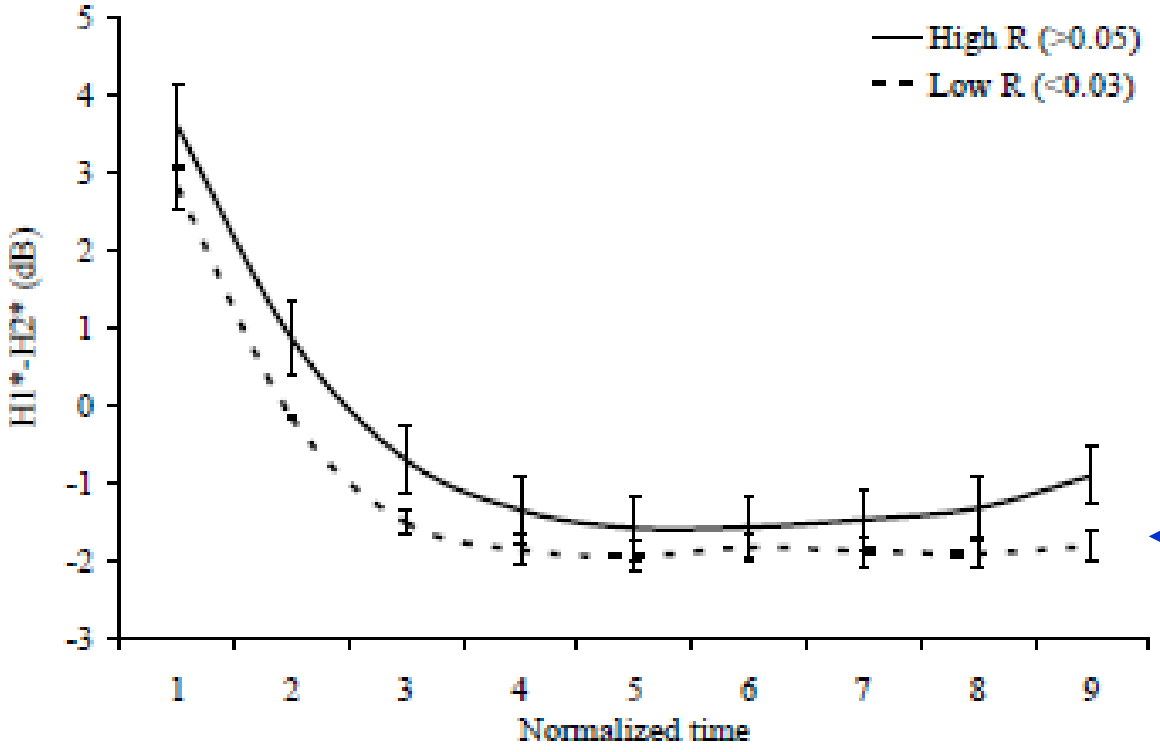
- In some languages, coda stops /p t k/ may be glottalized (with [ʔ])
- e.g. English words like *pat, got, hop, dock*
- (End of) preceding vowel will have smaller H1-H2, meaning it's creakier (dotted lines in graphs)





Garellek 2010, 2012

Lower-relative-frequency words show more coarticulation from /t/



creakier

Garellek 2011a

Listeners use this creaky voice to recognize English coda /t/

faster with creak

more accurate with creak

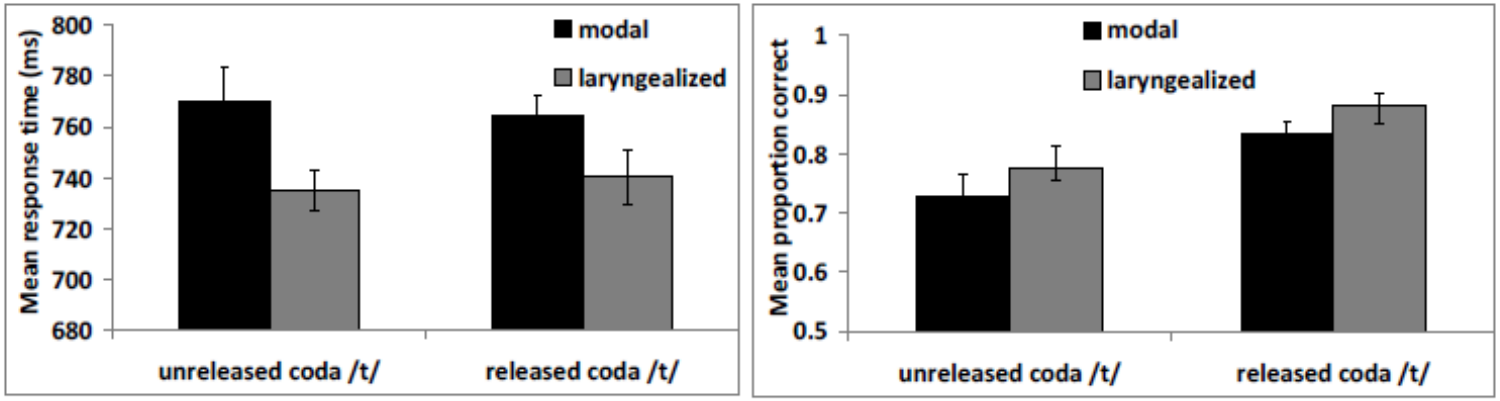


Figure 1. Mean response time and proportion correct of /t/ identification in targets. Error bars show standard error values.

Garellek 2011b

Sources of voice variation

- Variation in voice pitch
- Related prosodic variation
- Coarticulation from consonants
- Difficulties with voiced consonants
- Differences across individuals

Consonant voicing differences

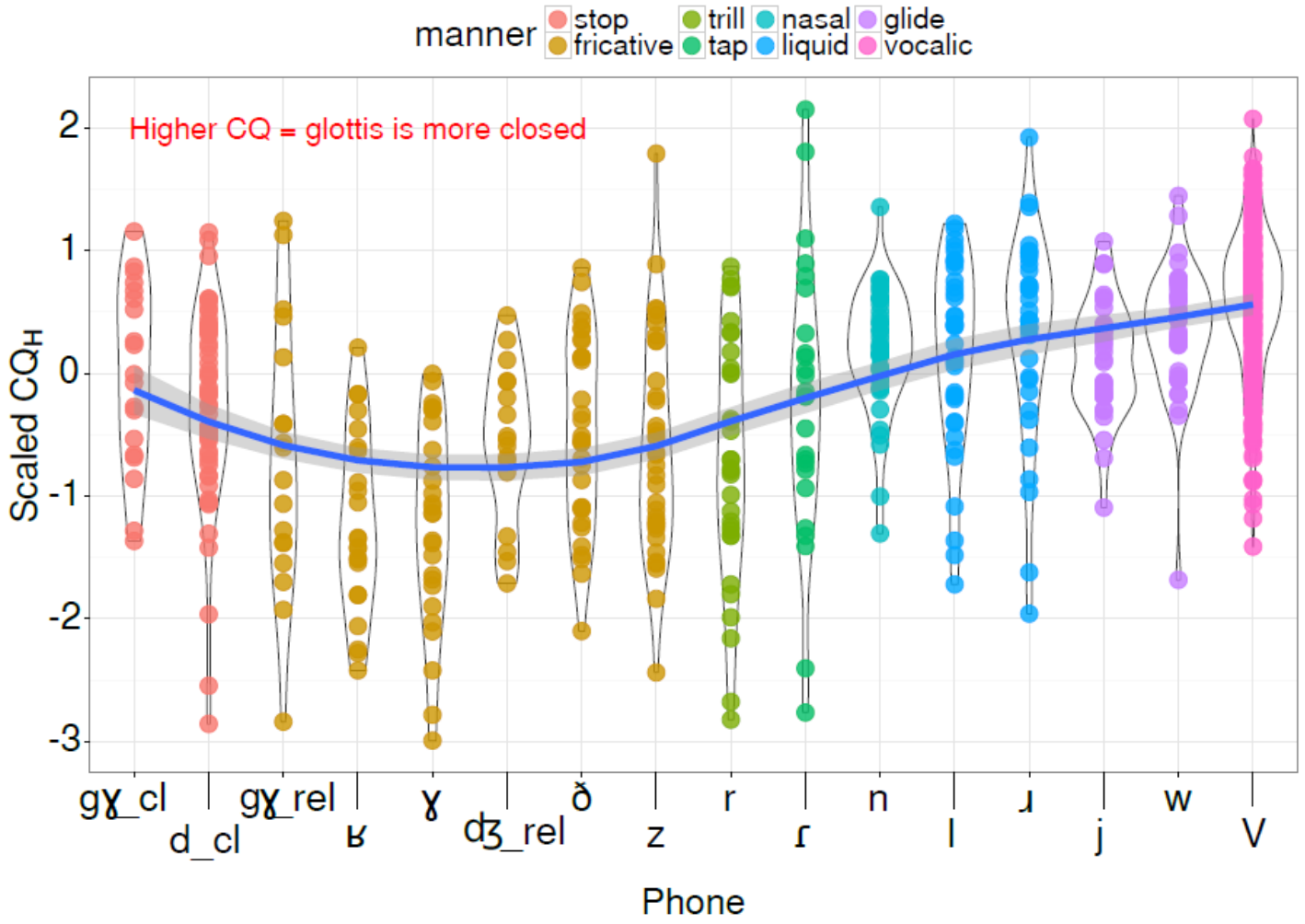
- Consonants differ in their oral constrictions and their airflow requirements
- Therefore must differ in difficulty of sustaining vocal fold vibration
- Can look at differences in vibration using electroglottography: voicing adjusts to the consonant

Consonant voicing: Does CQ differ across different consonants?

- EGG recordings of 14 speakers producing 14 consonants, 7 vowels; multiple reps
- Acoustically voiced constriction interval in each token (774 tokens)
- Mean Contact Quotient (CQ) for each interval
- (Standardized within speakers so speakers can be combined)

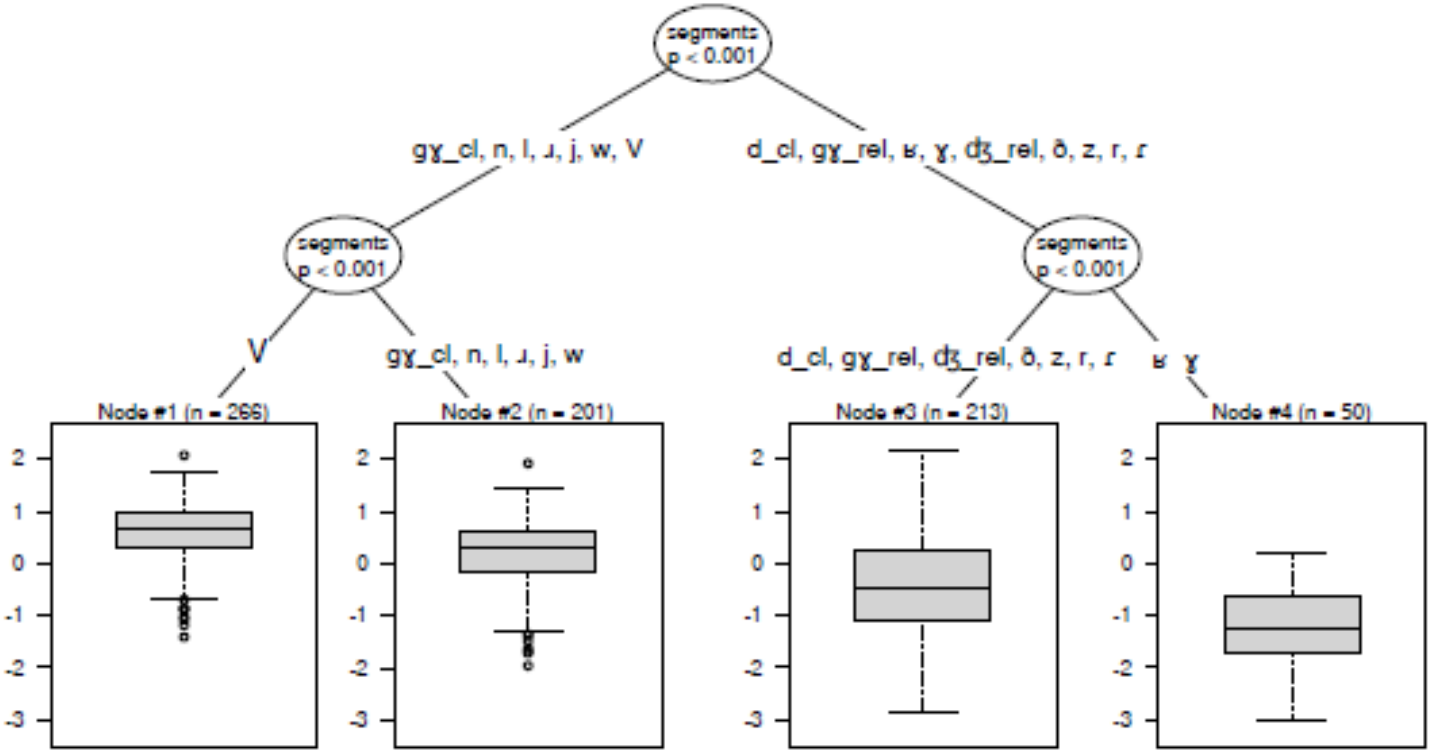
Risdal, Aly, Chong, Keating, Zymet (2016 CUNY and LabPhon)

CQ across all consonants + V



CQ_H: All Segments

- V > liquids/glides + [gɣ] (cl) > non-dorsal fricatives/tap/trill + [d] (cl) > dorsal fricatives



Sources of voice variation

- Variation in voice pitch
- Related prosodic variation
- Coarticulation from consonants
- Difficulties with voiced consonants
- Differences across individuals

Individual voice quality

- Voices differ in many ways – many acoustic properties characterize them
- We don’t yet know how important each acoustic property is to listeners when they recognize or distinguish voices
- General research strategy: compare the importance to voice perception of all measured acoustic properties

Individual voice quality:
How often do you sound more like someone else than like yourself?

- Corpus of voice samples from 200+ UCLA undergrads, each on 3 days and for multiple speech tasks
- Including reading 5 sentences 2x each x 3 days (= 30 sentences total per speaker)
- Perception experiment: For 3 speakers, listeners judged 2 non-identical tokens as *same speaker/different speakers*

Kreiman, Sigouin et al. in prep

Sample results

Reference speaker for this example

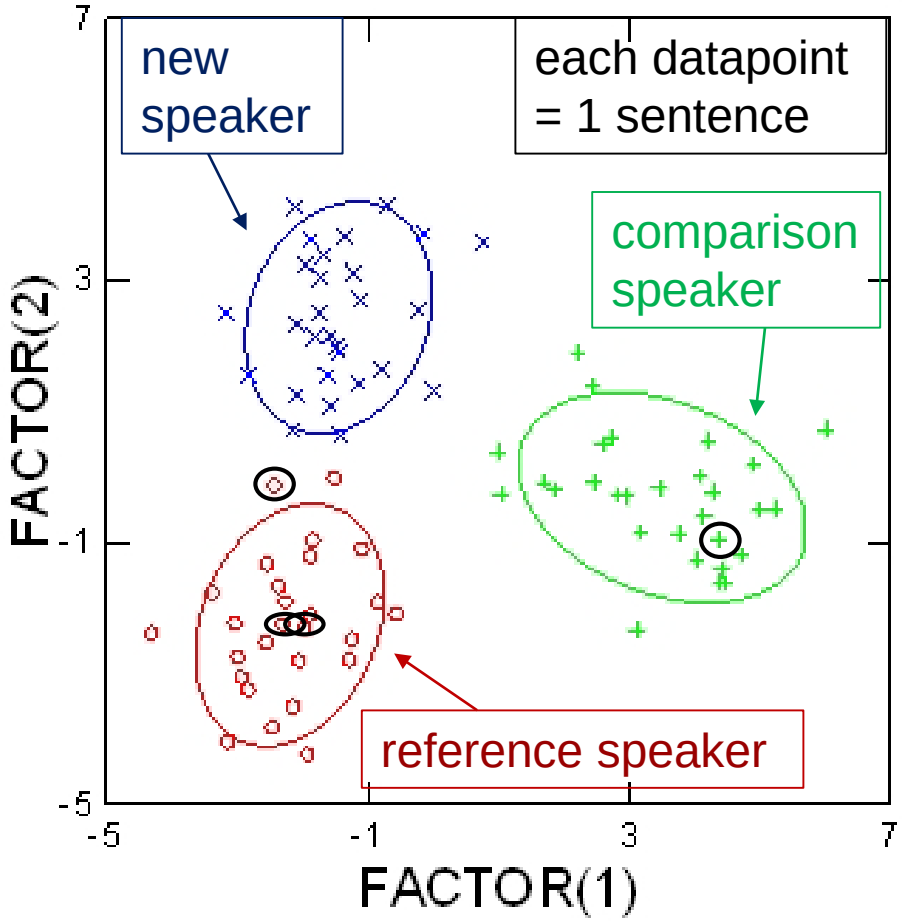
	Sounded like ONE speaker to listeners	Sounded like TWO speakers to (some) listeners
2 tokens produced by ONE speaker (the reference speaker)	(100% correct)	(67% correct)
2 tokens produced by TWO speakers (reference speaker and comparison speaker)		(100% correct)

How much do the three voices differ acoustically?

- Acoustic measures of all vowels + sonorants from each sentence: sentence mean/SD for each measure
- Select the uncorrelated measures
- Linear Discriminant Analysis of data labeled for speaker
- LDA classifies all tokens correctly for the speakers, with only 2 dimensions

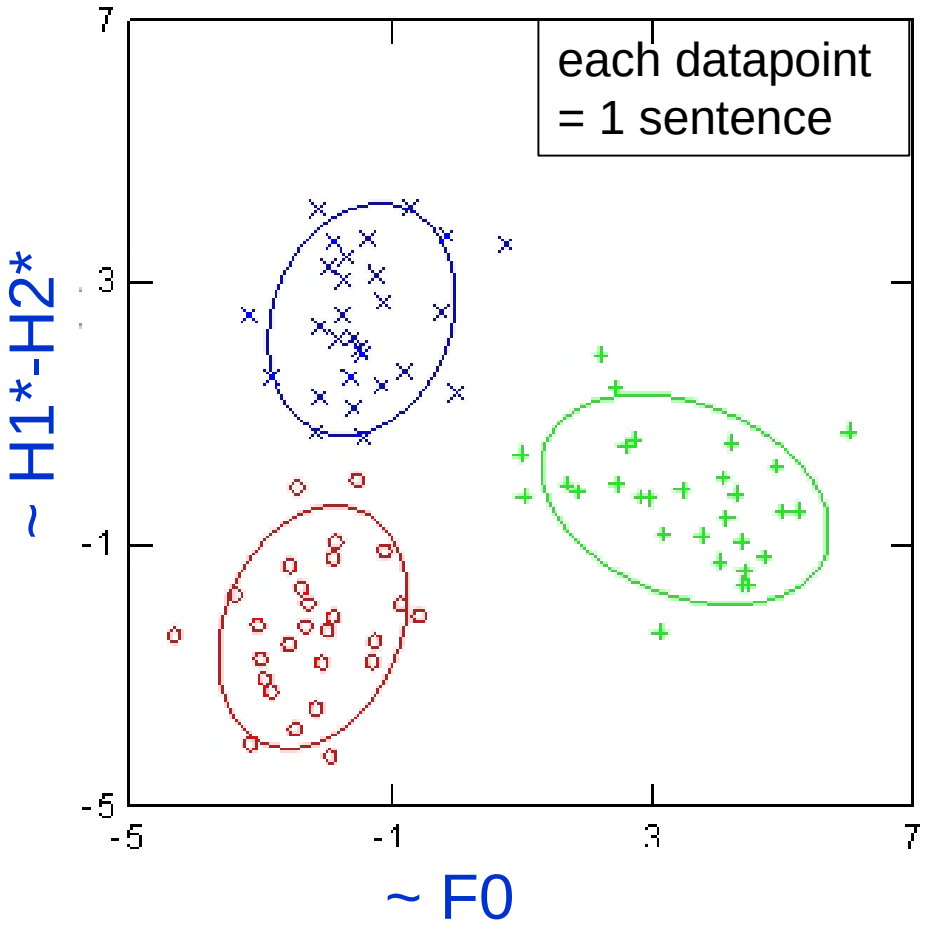
Voice discrimination space

- In this 2-D space defined by the discriminant function, all tokens of the 3 speakers are distinct



Voice discrimination space

- LDA Factor 1 (x) ~ F0
- separates green from others
- LDA Factor 2 (y) ~ H1*-H2*
- separates red from blue



Conclusions

- Even in a language like English, without phonation contrasts on consonants or vowels, voice quality varies constantly throughout connected speech because of segmental and prosodic context
- In effect each speaker’s voice is a distribution of qualities, potentially overlapping with other voices

Further acknowledgments

- NSF grants BCS-0720304, IIS-1018863, IIS-140992
- Students in Winter 2015 Speech Production course, for segment expt.
- Linguistics undergrads labeling the multi-speaker speech corpus
- Ann Aly for figures