Voice quality variation within and across languages

Pat Keating

UCLA Linguistics Department

Linguistic Society of America Annual Meeting
Jan. 2016, Washington DC

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)

back

front

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 ways to measure voice variation
- Cross-language phonation contrasts
- Voice quality and pitch: dependence and independence
- Differences across consonants
- Differences across individuals

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

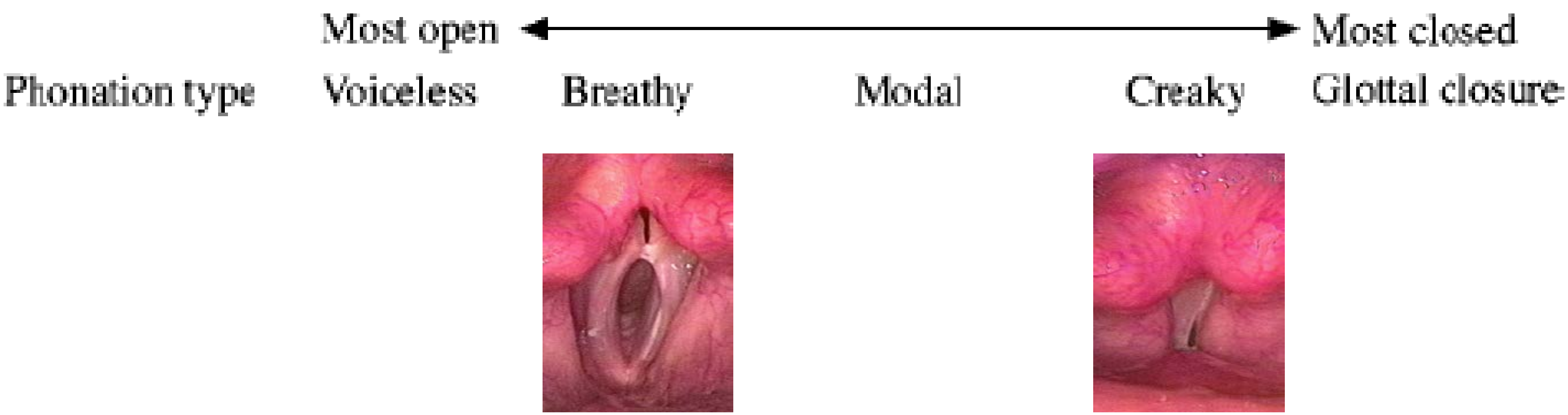
Yen-Liang Shue
Dolby Australia

Caroline Sigouin
U Laval

Phonation contrasts in languages of the world

- Many languages contrast phonations on vowels and/or consonants
- Common especially in SE Asia, the Americas, India

Ladefoged’s glottal continuum



IPA diacritics: a̤ a̰

On the breathy side of modal: lax, slack, or lenis
On the creaky side of modal: tense, stiff, fortis, or pressed

Ladefoged (1971) *Preliminaries to linguistic phonetics*

New tools for voice analysis

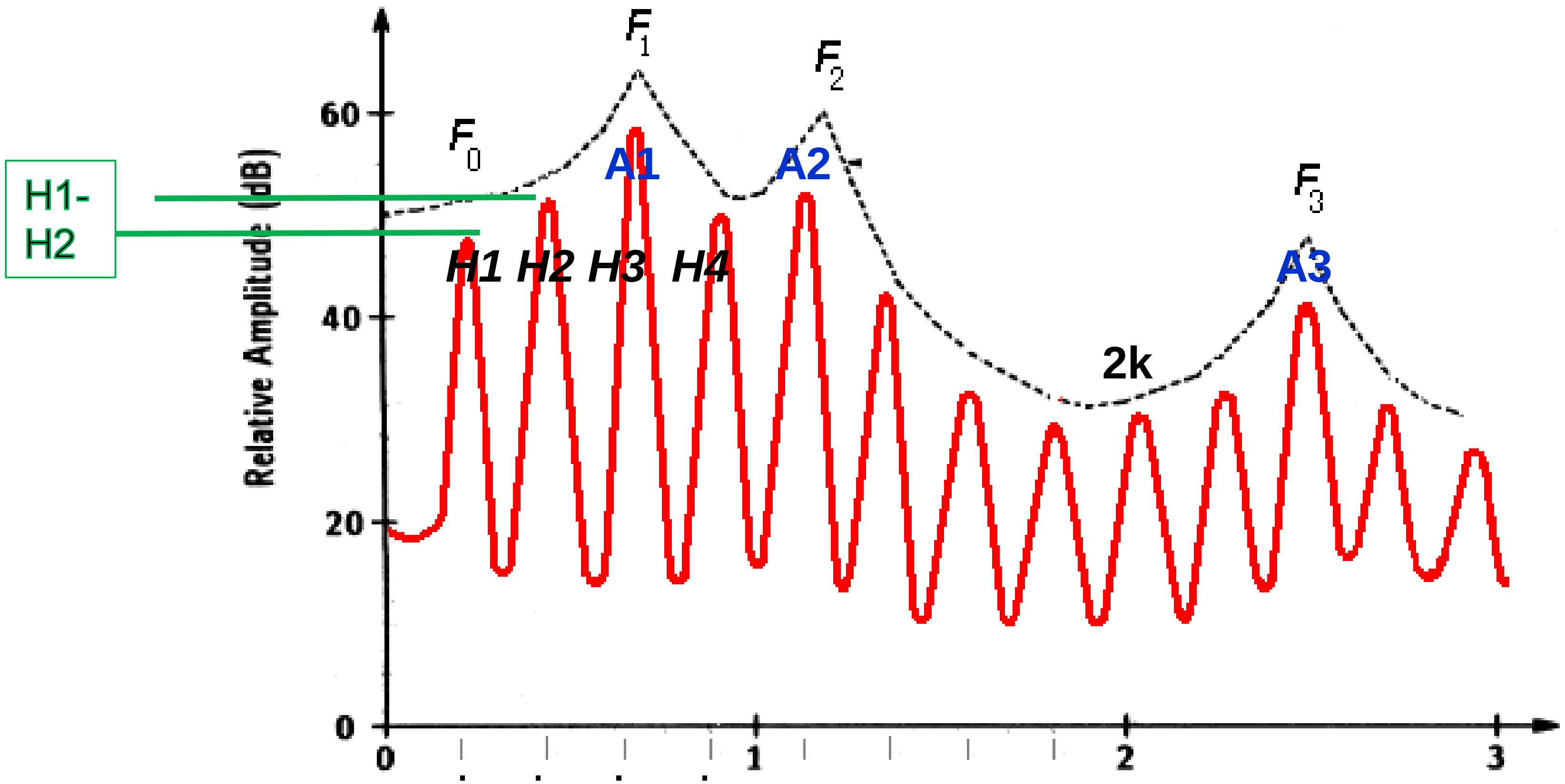
- For acoustic analysis: VoiceSauce

- For physiological analysis: EggWorks, used with VoiceSauce

- Both = UCLA free software

Shue 2010, Shue et al. 2011, Tehrani 2012

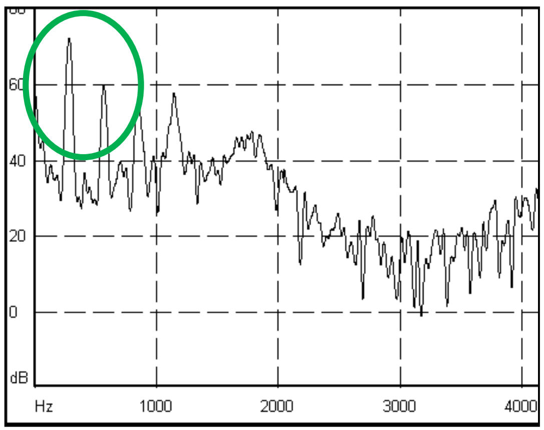
Acoustic measures based on harmonics in spectrum



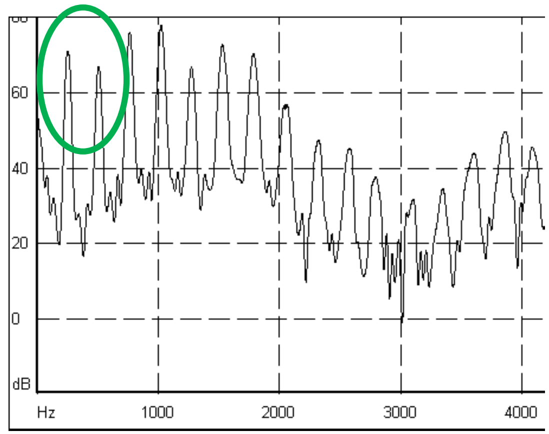
Shue 2010, Kreiman et al. 2007, Kreiman et al 2014

H1-H2 example: Jalapa Mazatec

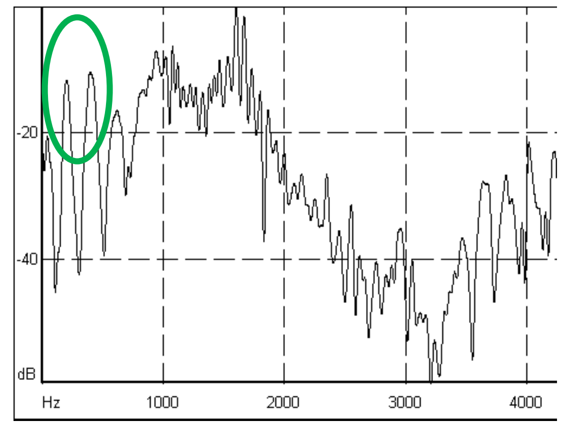
breathy



modal



creaky



Breathy	Modal	Creaky
ba̤^{34}	ba^{32}	ba̰^{3}

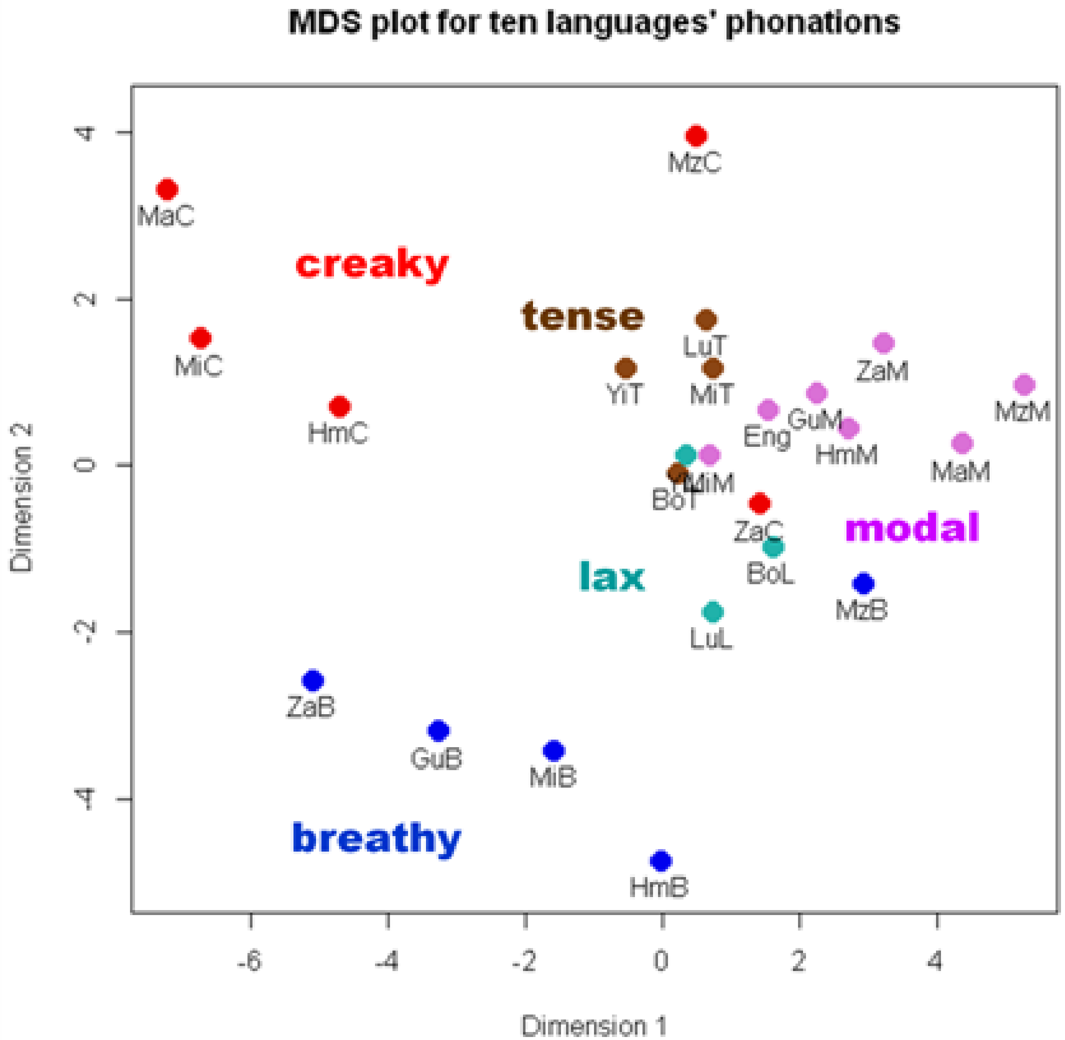
Kirk et al. 1984, Garellek & Keating 2011

Acoustics of vowel phonation contrasts across languages

- Recordings from 10 languages, men only
- Coded for 5 phonation categories – Breathy/Lax/Modal/Tense/Creaky
- 24 instances (some contrastive, some not), with means on all acoustic measures
- Multi-Dimensional Scaling to reduce high-dimensional data to a low-dimensional map of acoustic distances

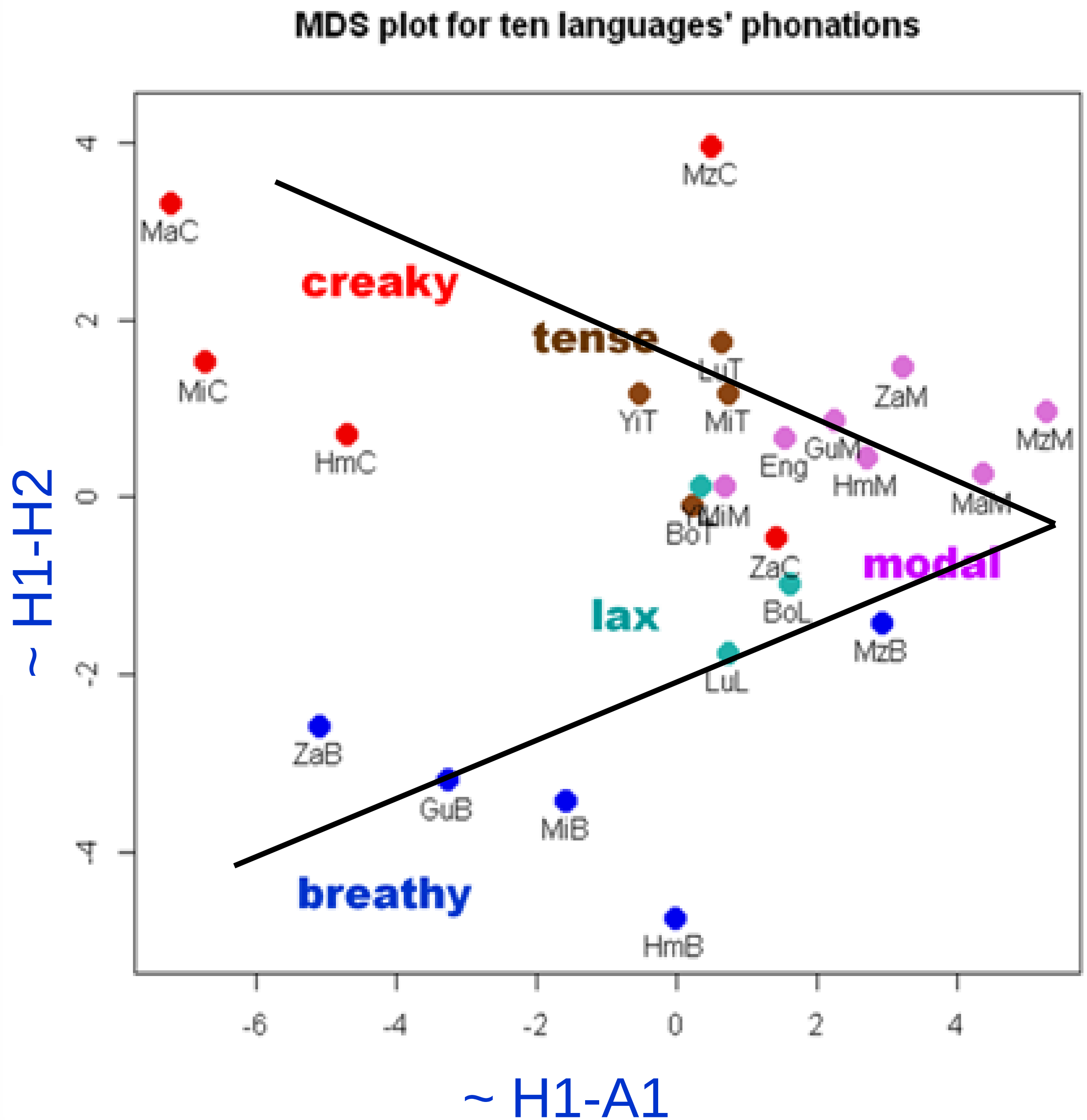
2-D acoustic space from MDS

Languages:
Bo
English
Gujarati
Luchun Hani
Hmong
Mandarin
Mazate**c**
Miao (Black)
Yi (Southern)
Zapotec (Valley)



Keating et al. 2012

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



Keating et al. 2012

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

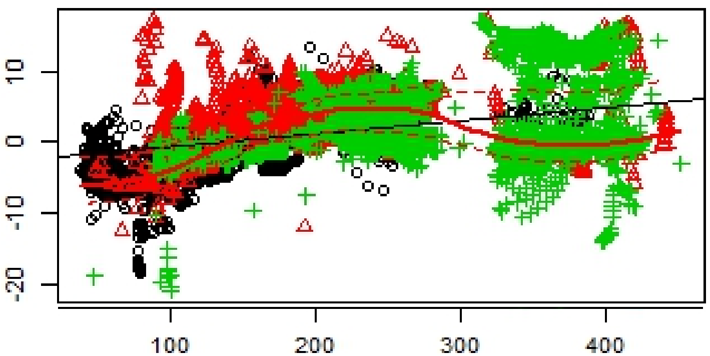
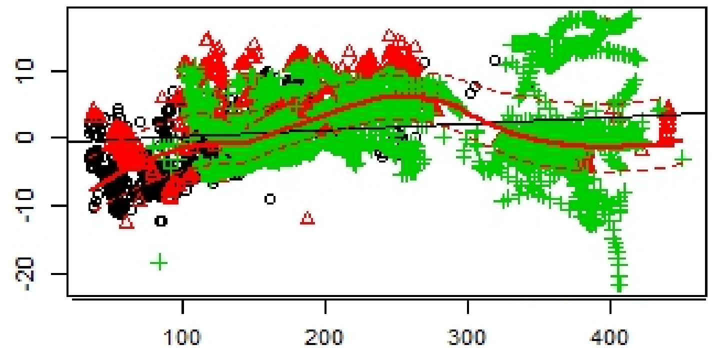
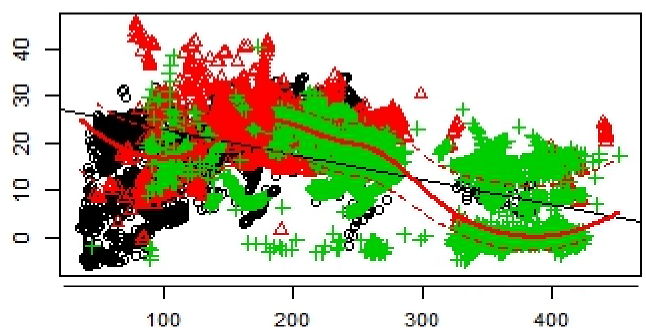
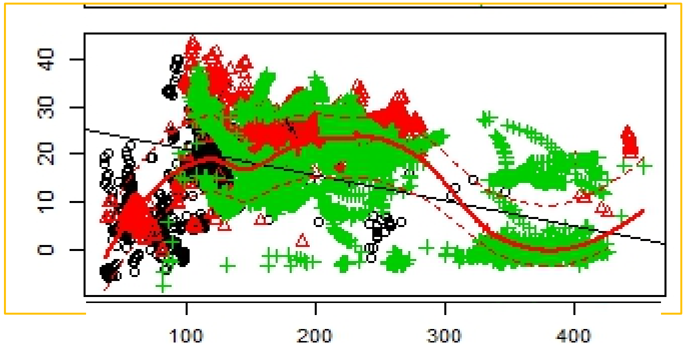
Voice quality in relation to voice pitch

- 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

How does this work?

- Audio recordings of pitch glides up or down by English and Mandarin men and women
- 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

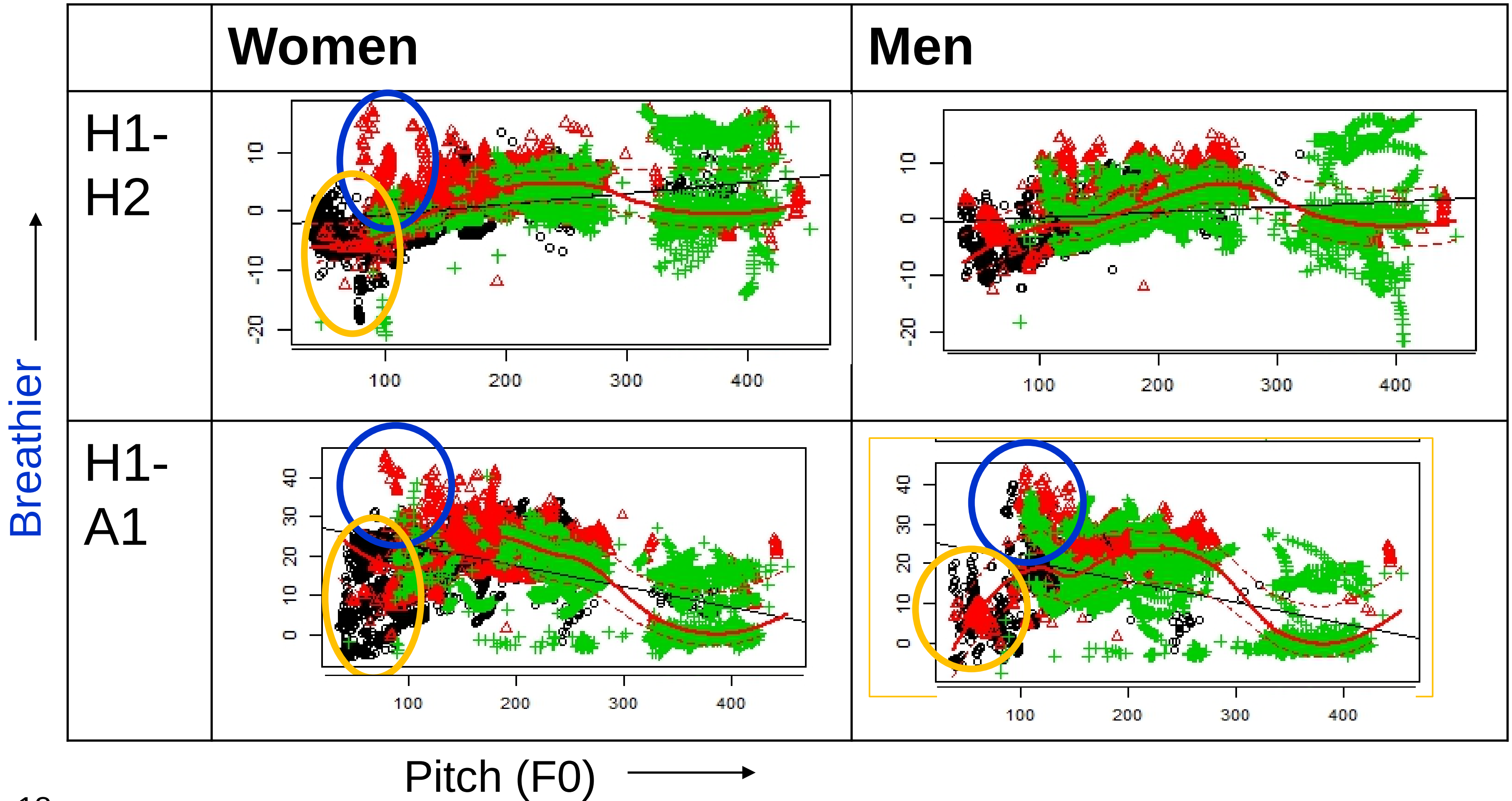
	Women	Men
H1-H2		
H1-A1		

Breathier

Pitch (F0)

Red △ = falling pitch (don't creak)
Black ○ = falling pitch (creak is ok)
Time runs right-to-left

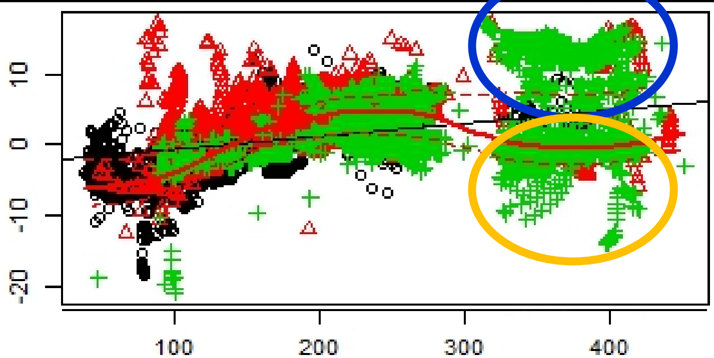
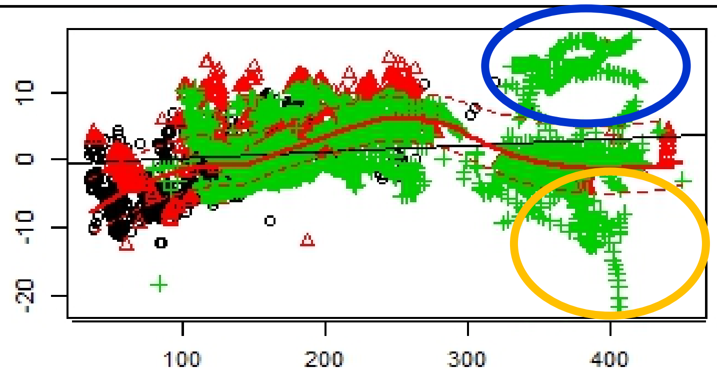
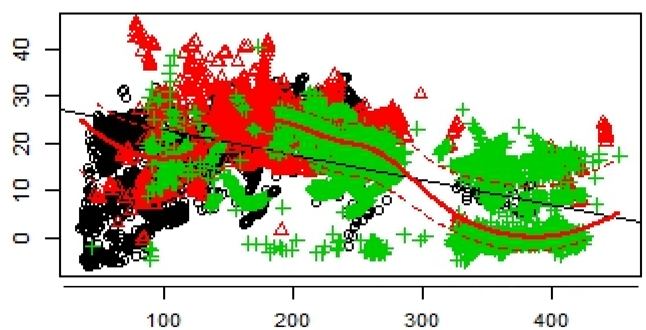
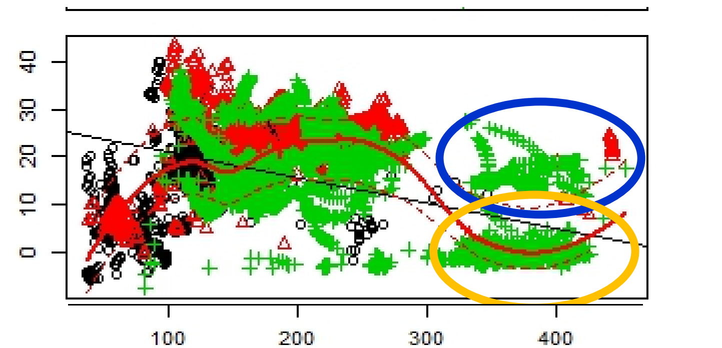
Green + = rising pitch
Time runs left-to-right



Red ∆ = falling pitch (don't creak)

Black ○ = falling pitch (creak is ok)

Green + = rising pitch

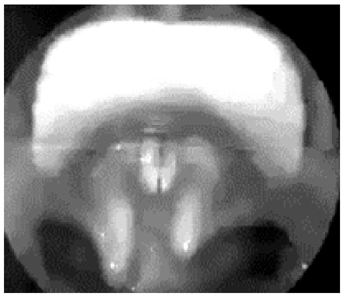
	Women	**Men**
H1-H2		
H1-A1		

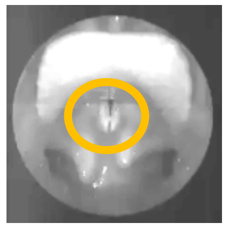
Breathier

Pitch (F0)

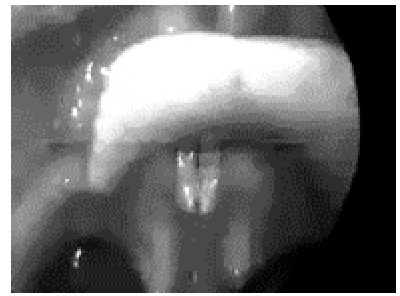
Two strategies to raise pitch:

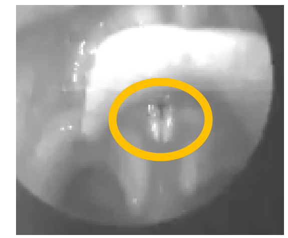
Up to falsetto voice





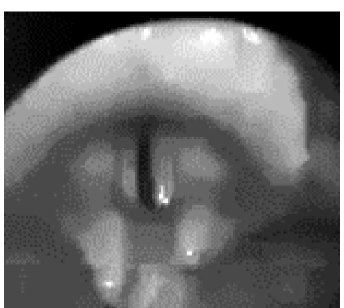
Up to tense voice

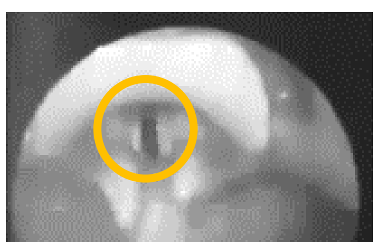




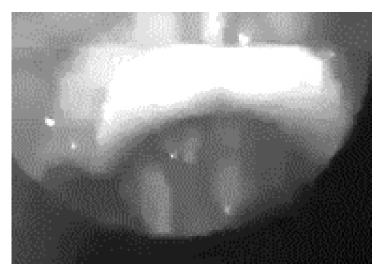
Two strategies to lower pitch:

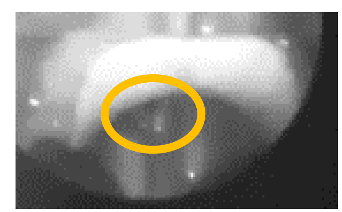
Down to breathy voice





Down to creaky voice





Thanks to Jody Kreiman, Bruce Gerratt, and Henry Tehrani for help making these video clips

Phonation contrast on low tones: Santa Ana del Valle Zapotec

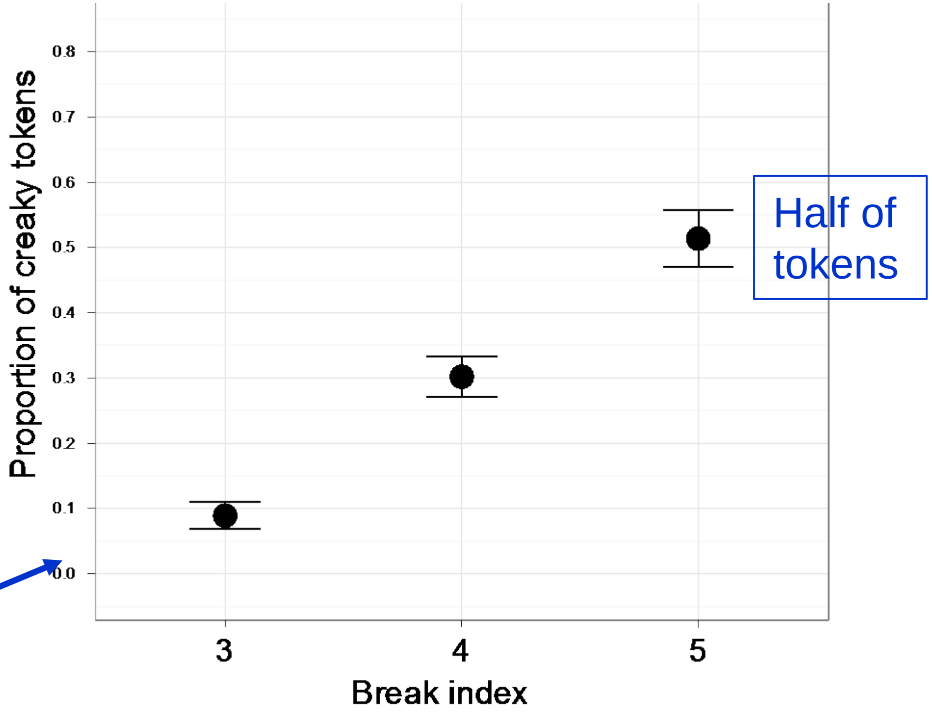
- Modal High (and Rising) tones
- Breathy and creaky Falling tones
- Contrasts with Modal-High:
  - Breathy: ‘place’ **la̤t**
  - Creaky: ‘field’ **la̰ts**
  - Modal: ‘can’ **lat**

Esposito 2010

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

Incidence of creaky voice by Break Index



Small Phrase | Full Phrase | Utterance

Garellek & Keating, 2015 LSA

Especially important in tone languages

- Pitch-range expansion is crucial for tone systems with more than 2 minimal pitch contrasts (2 levels / 2 rises / 2 falls)
- So lowest tones tend to be creaky- or breathy-voiced
- And highest tones tend to be tense-voiced, or even falsetto

Kuang 2013a,b

But tone/phonation languages can break the correlation

- Tone and phonation contrasts can be independent, combining orthogonally within a single language
- In these languages, speakers must largely de-couple 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?

Cross-classifying example: Mpi (plays by rows)

TONE (PITCH)	REGULAR VOICE	ENGLISH	TENSE VOICE	ENGLISH
Low rising	**si**	‘to be putrid’	**si**	‘to be dried up’
Low level	**si**	‘blood’	**si**	‘seven’
Mid rising	**si**	‘to roll rope’	**si**	‘to smoke’
Mid level	**si**	(a color)	**si**	(classifier)
High falling	**si**	‘to die’	**si**	(name)
High level	**si**	‘four’	**si**	(name)

James Harris & Peter Ladefoged: http://www.phonetics.ucla.edu/vowels/chapter12/mpi.html

Yi languages: cross-classifying *tense* vs *lax* with tones

	Low tone	**Mid tone**
Lax phonation	be ^{21} (mountain)	be ^{33} (fight)
Tense phonation	be ^{21} (foot)	be ^{33} (shoot)

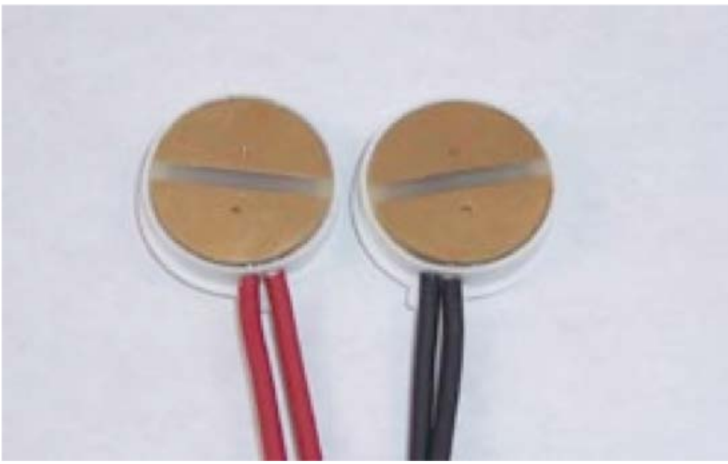
Example from Southern Yi

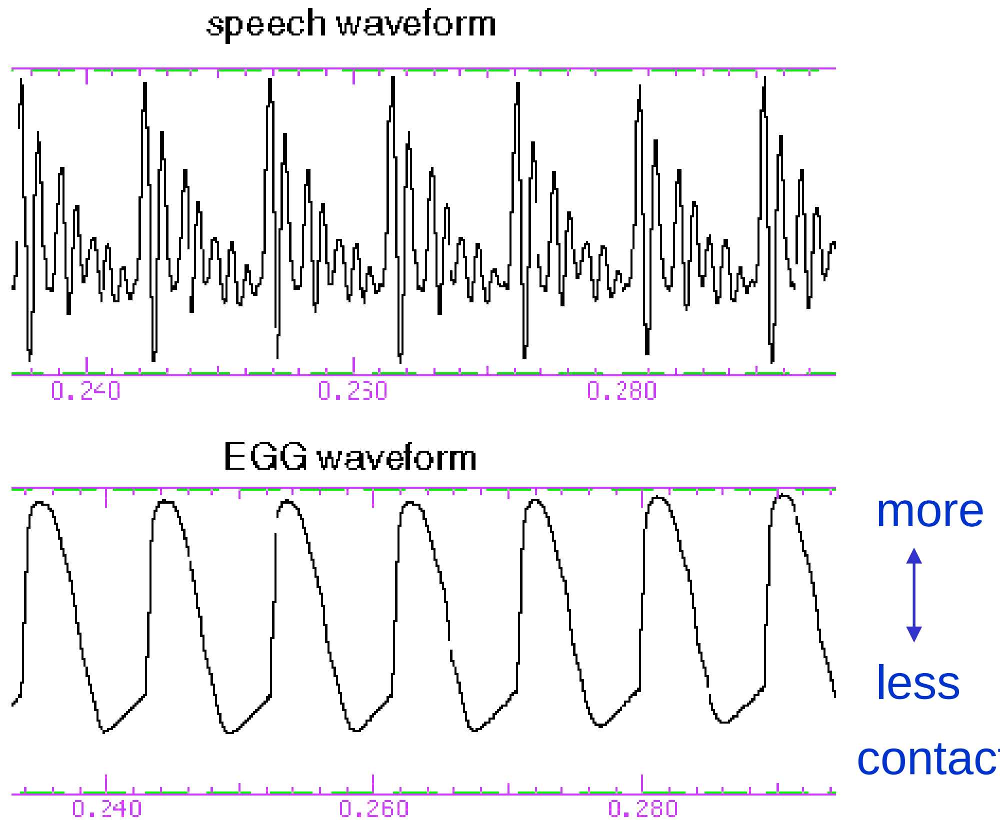
Kuang 2011, Kuang & Keating 2014

Fieldwork in Yunnan (J. Kuang)

Kuang 2011, 2013

Electroglottography (EGG)





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 voice)

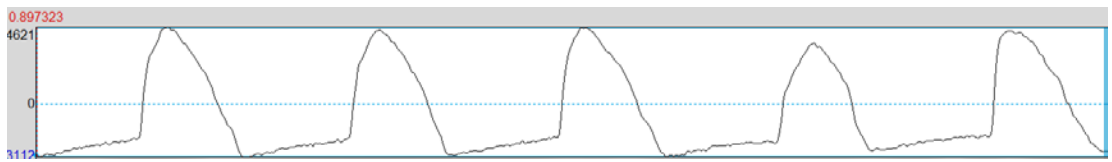
Rothenberg & Mahshie 1988, Herbst & Ternstrom 2006

CQ example: White Hmong

EGG waveforms of 3 phonations

Breathy:

CQ = .41

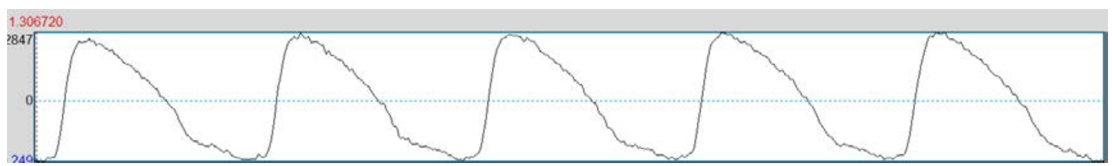


more
contact

less
contact

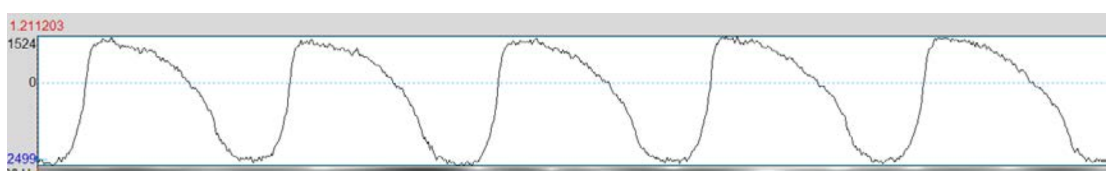
Modal:

CQ = .57



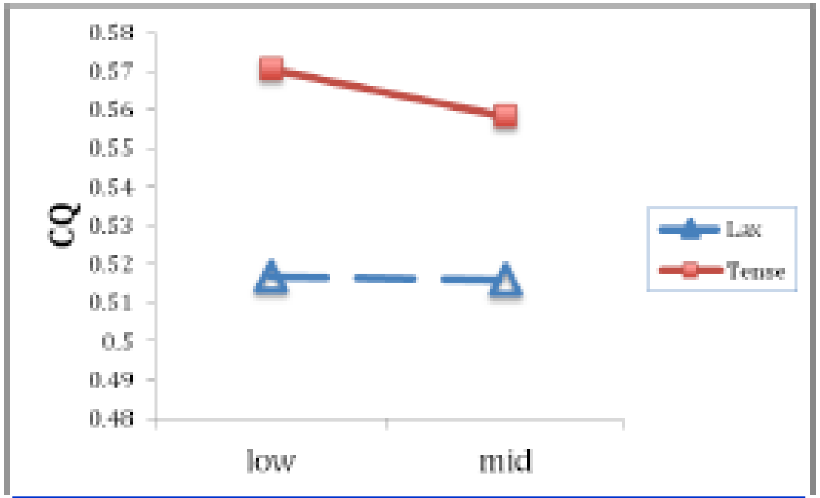
Creaky:

CQ = .65



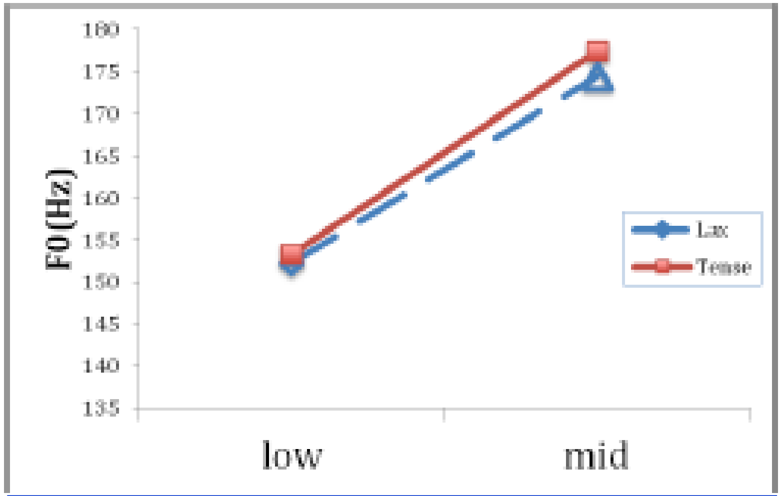
Tense vs *lax* in 3 Yi languages: low vs. mid tones

No tone effect on CQ



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

No phonation effect on F0



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

Kuang & Keating 2012, Kuang 2013

Summary, voice quality and pitch

- Voice quality generally varies with voice pitch, allowing pitch-range expansion
  - For intonation
  - For lexical tones
- 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)

Consonant voicing differences

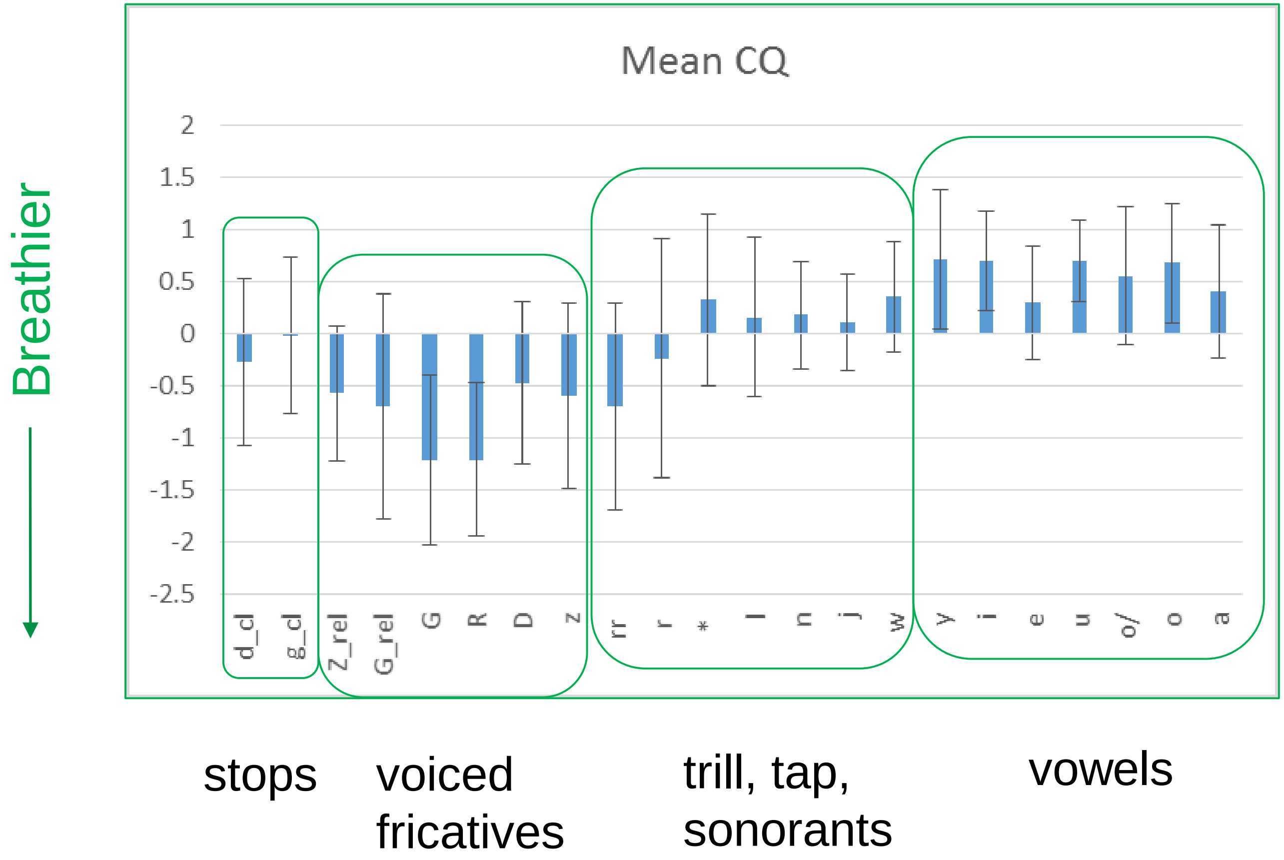
- Consonants differ in their oral constrictions and their airflow requirements
- Therefore must differ in difficulty of sustaining vocal fold vibration
- Can look at differences in vibration using electroglottography

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 for each interval
- (Standardized within speakers so speakers can be combined)

Risdal, Aly, Chong, Keating, Zymet
(2016 CUNY and in prep)

Mean standardized CQ for 21 segment types



Risdal et al. in prep

Summary, cross-segment voicing

- It’s not just speakers of languages with phonation contrasts who produce variation in voicing
- Speakers of other languages vary details of how the vocal folds vibrate in order to facilitate voicing – here, across a variety of segment types
- Presumably an example of ease of articulation

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

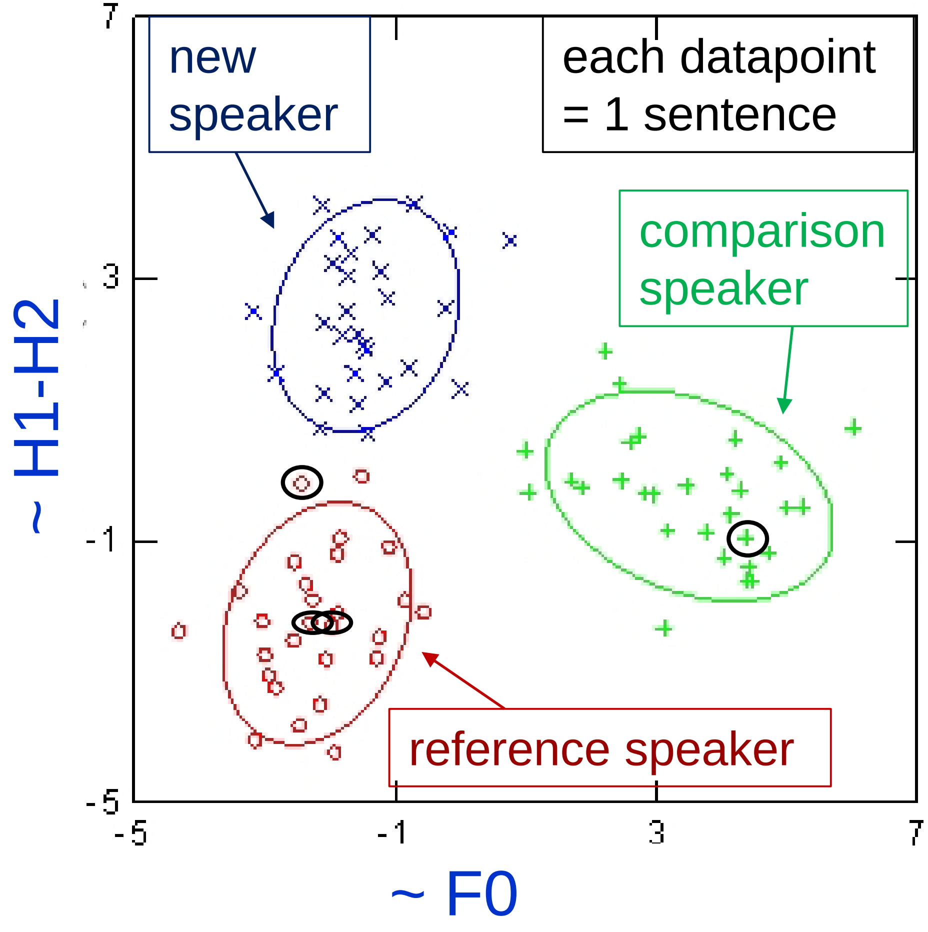
Sample results

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 vowels + sonorants: sentence means/SDs
- 2-D acoustic space for all 90 sentences
- Factor 1 (x) ~ F0
- separates green from others
- Factor 2 (y) ~ H1-H2
- separates red from blue



Next step: Can we predict perception from acoustics?

- Not yet! In this experiment, these 2 factors don’t predict perception in detail, beyond the fact that F0 is very important
- Most likely our current acoustic measures aren’t complete, and their weighting in perception is complex
- Work in Jody Kreiman’s lab on a perceptual model of the voice spectrum will be important

Summary, cross-speaker

- A toy problem: distinguishing 3 voices
- Requires (at least) 2 acoustic dimensions
- These dimensions are very familiar: F0 and H1-H2, both important in linguistic contrasts
- Future work will explore what other acoustic dimensions structure voice perception

Overall summary -1

- Speakers have identifiable personal voices
- Speakers vary their voice quality
  - Because they speak a language with a phonation contrast
  - Because it makes it easier to vary pitch
    - In a tone system
    - In an intonational system
  - Because some segments are harder to voice

Overall summary - 2

- The variation across phonation categories may lie in a 2-D acoustic phonetic space
- Defined by the harmonic spectrum from the fundamental harmonic up to ~ F1
- These acoustic dimensions also relevant for pitch-phonation relations and individual voice quality

Questions

- Is this phonetic space valid for more languages and contrasts?
- How does it fit with models of voice production in general?
- How would a single speaker’s variation due to pitch and consonant type fit in?
- How are individual voices located in it?

Conclusions - 1

- For phonetics, we have a hypothesis about two important parts of the harmonic spectrum – a hypothesis in accord with Kreiman’s perceptual model of voicing
- For everyone else, we have new tools to facilitate the study of voice from different perspectives

Conclusions - 2

Voice quality variation is a relatively new, but rich, research topic, not only for phonetics and phonology, but for any area where personal identity, as expressed by the voice, is relevant. We hope that the new tools for voice research will encourage non-phoneticians to include voice analysis in their research toolbox.

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
- JJ Kuang for data analysis; Ann Aly for figures