Voice quality measurement and error correction

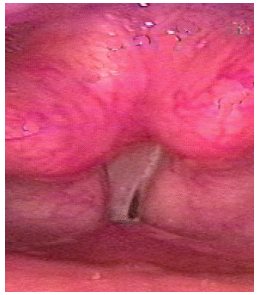
Yuan Chai

Phonetics lab meeting

05/31/2024

Ladefoged's glottal continuum

Phonation type	Most open → Most closed				
	Voiceless	Breathy	Modal	Creaky	Glottal closure



IPA diacritics: a̤ a̰

Ladefoged (1971) *Preliminaries to linguistic phonetics*

Slide from Patricia Keating

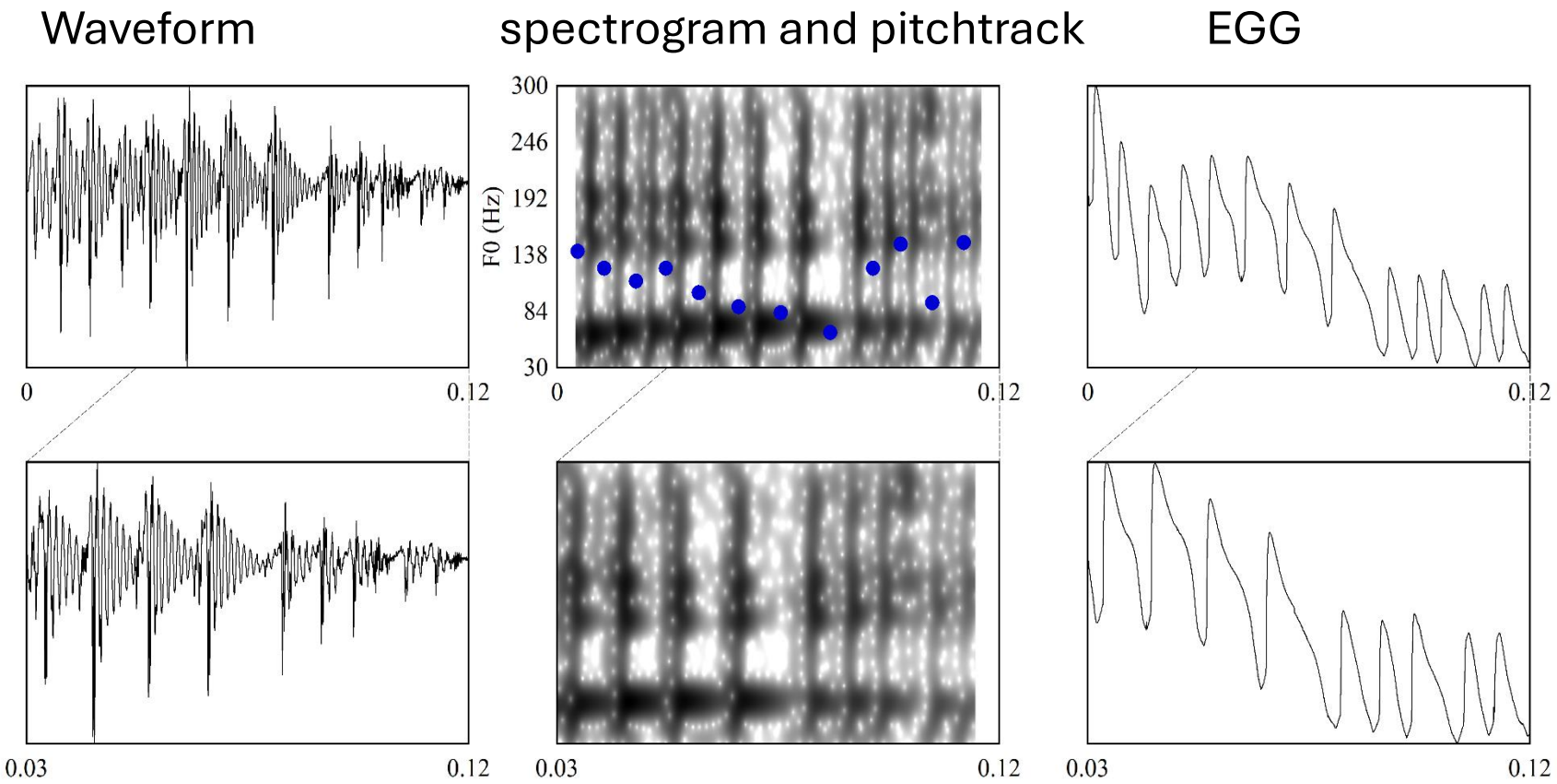
Prototypical creaky voice

- Keating et al. 2015, Keating et al. 2023
- Constricted glottis
  - Acoustic parameter: H1–H2 (?) or Residual H1 (?), Contact quotient
- Irregular vocal folds vibration
  - Acoustic parameter: Harmonic-to-Noise ratio (HNR), Cepstral Peak Prominence (CPP)
- Low f0

- Prototypical creaky voice is not as “prototypical” as we thought
- Actually quite difficult to find a token that has all three parameters

Prototypical creaky voice

- Lowest F0: 64 Hz; Mean CQ: 0.61



Prototypical creaky voice

- Problems with measuring the acoustic measures
  - Because the f0 is irregular, very likely to get f0 tracking errors from the tracking algorithm.
  - If f0 value is wrong, the amplitude of H1 and H2 is definitely measured at the wrong frequency.
- Solution:
  - f0: manually correct f0
  - H1–H2: if there is very little periodicity, it is impossible to have a spectral analysis
  - Use CQ value to represent the degree of constriction.

Prototypical creaky voice

- In the following slides, we are going to explore:
  - Whether these three parameters can stand alone and create a creaky percept
  - Whether the combination of every two parameters can yield a creaky percept

Low f0 creak

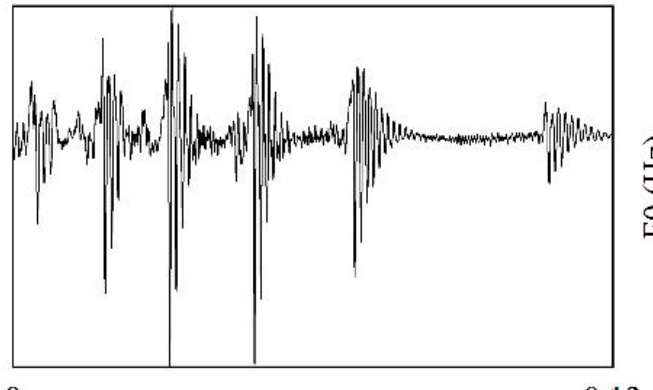
- low f0
- regular periodicity
- modal vocal folds constriction

- This is what usually called “vocal fry”.
  - The term “vocal fry” should not be used equivalently as “creaky voice”, since it is a subtype of creaky voice.

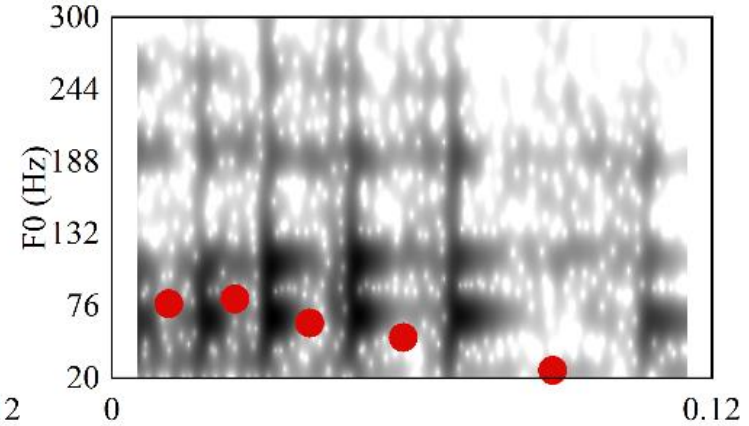
Low f0 creak

(F0: 26-81 Hz; CQ: 0.31, spread glottis)

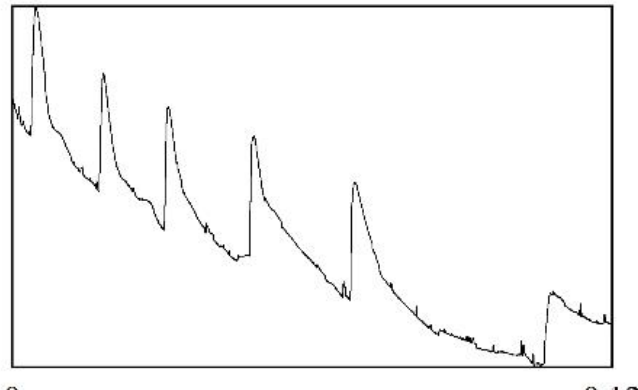
Waveform



spectrogram and pitchtrack



EGG



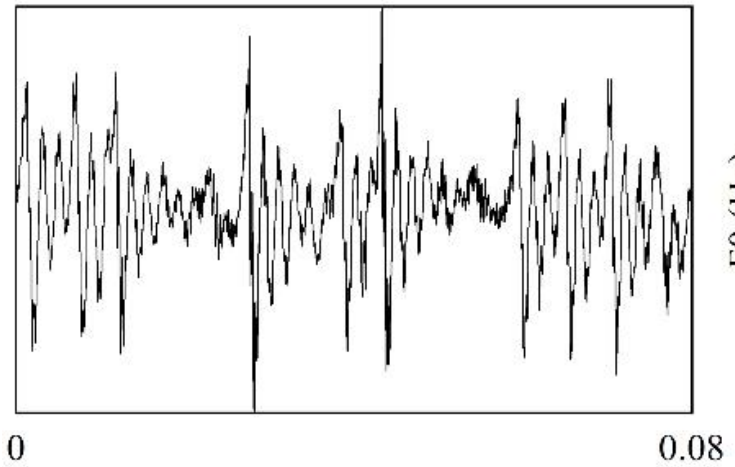
Irregular f0 creaky

- involve non-low f0
- irregular f0
- modal vocal fold constriction

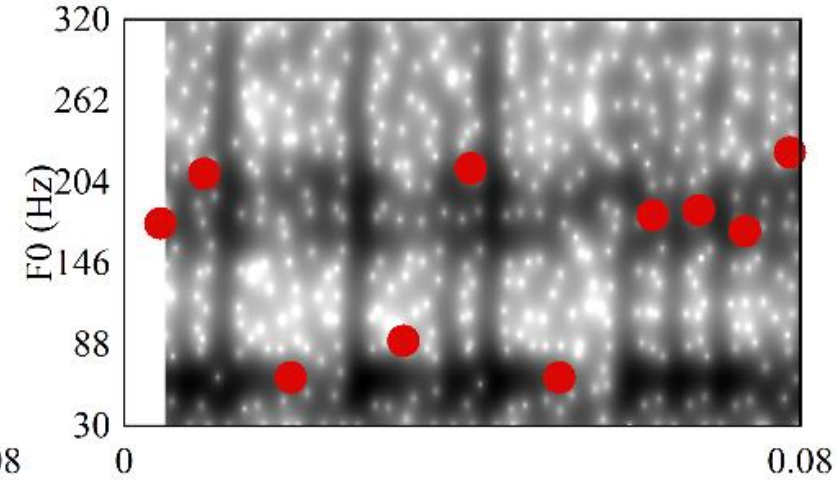
Irregular f0 creaky

(F0: 62-246 Hz, mix of random and period-doubled; CQ: 0.38, spread glottis)

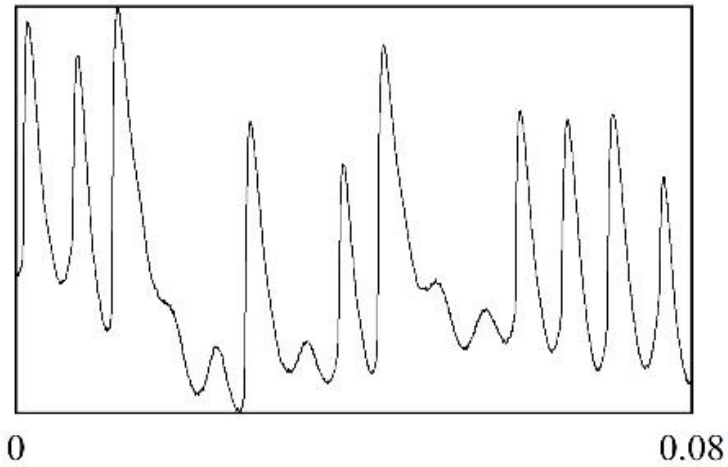
Waveform



spectrogram and pitchtrack



EGG



Constricted glottis (not creak)

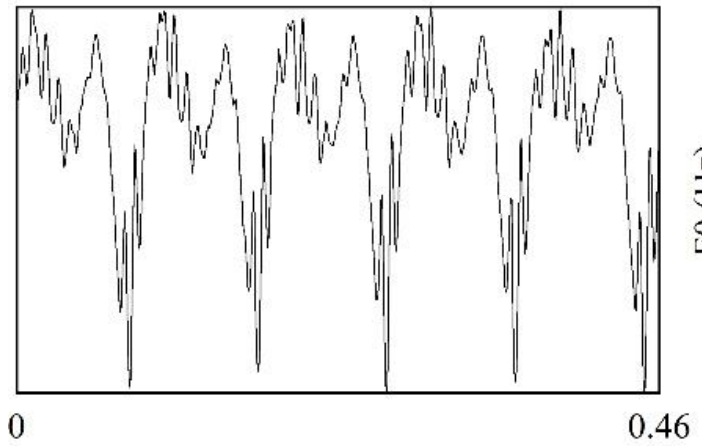
- non-low f0
- regular f0
- constricted vocal folds

- This is not creak. It generates tense voice.

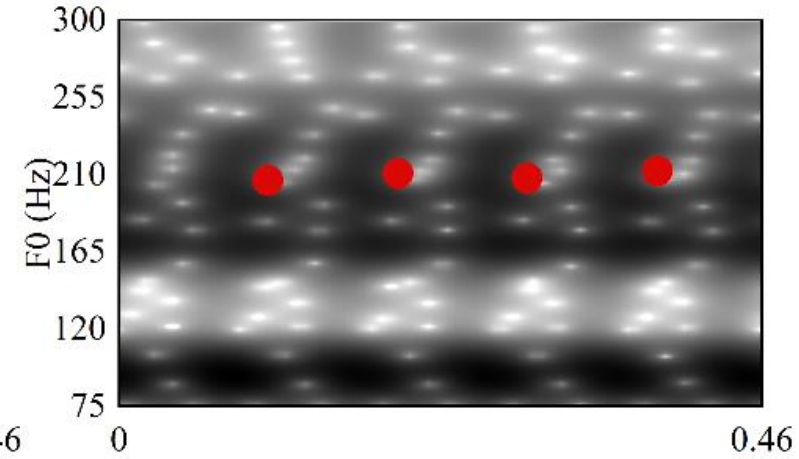
Constricted glottis (not creak)

(F0: 211 Hz; CQ: 0.61, constricted)

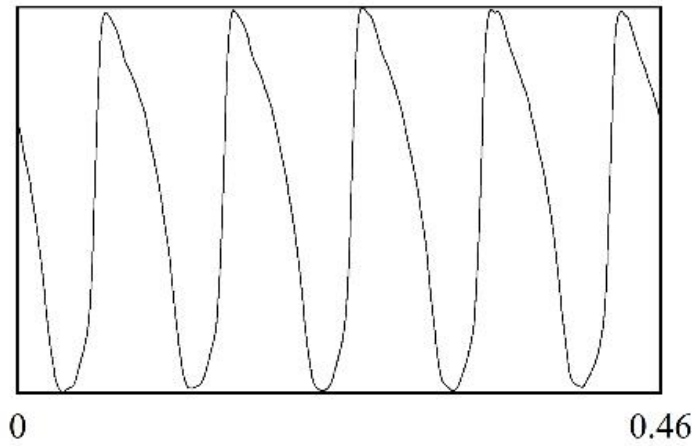
Waveform



spectrogram and pitchtrack



EGG

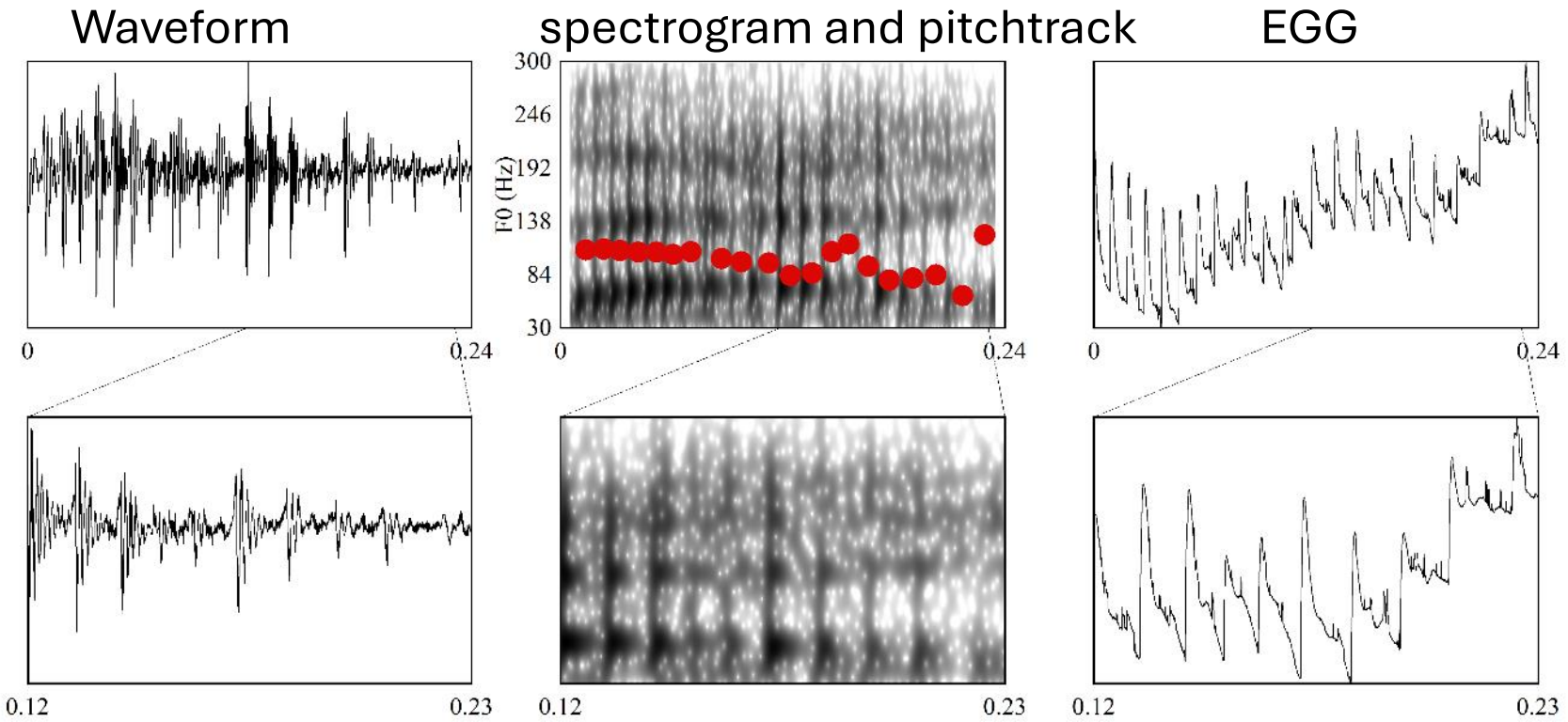


Low f0 + Irregular f0 creak

- low f0
- irregular f0
- normal vocal fold constriction

Low f0 + Irregular f0 creak

(F0: 62-124 Hz; CQ: 0.34, spread glottis)

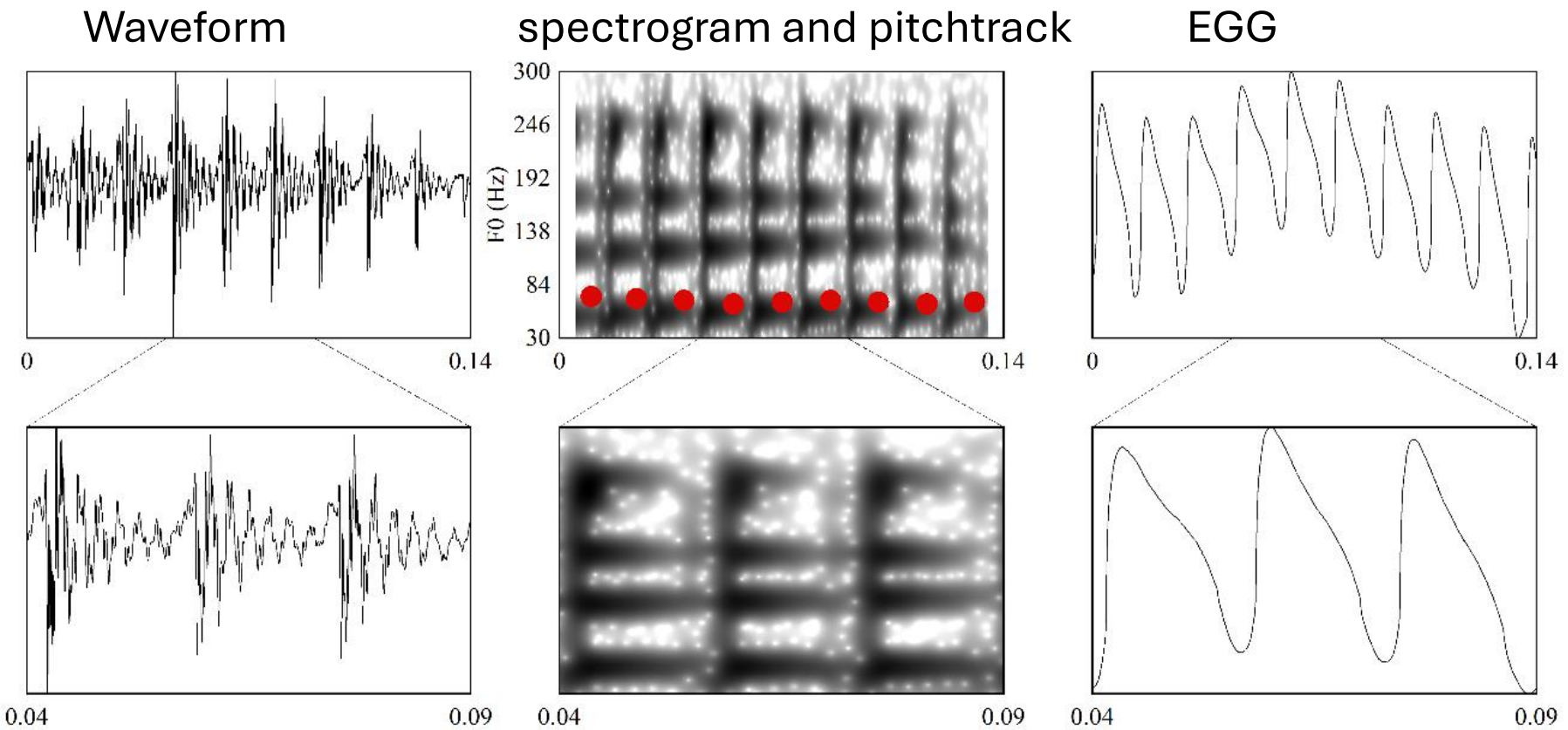


Low f0 + Constricted glottis creak

- low f0
- regular f0
- constricted glottis

Low f0 + Constricted glottis creak

(F0: 67 Hz; CQ: 0.63, constricted glottis)



Irregular f0 + Constricted glottis creak

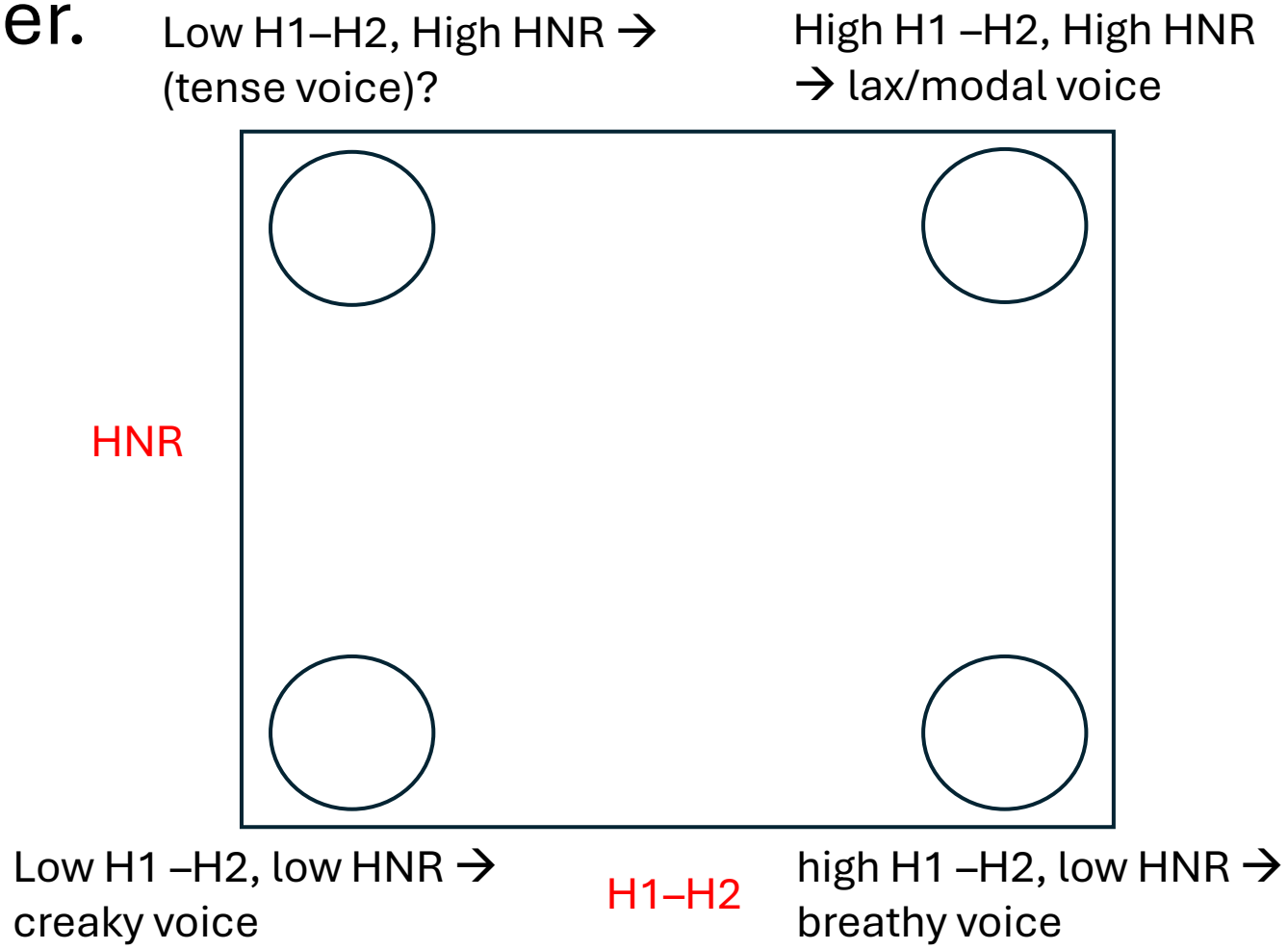
- Haven’t found any token fitting in this category (yet)

Interim summary

Type – sounds creaky	Low F0		Low HNR	High SHR	Low H1–H2	High CQ
Low F0	√			(optional)		
Irregular F0			√		(not defined)	
Prototypical creak	√		√		√	√
Vocal fry	√				√	√
Spread glottis creak	√	(and/or)	√		NO, high	
Multiple-pulsed spread glottis creak	√	(and/or)	√	√	NO, high	
Type – does not sound creaky						
Tense voice					√	√
Multiple-pulsed				√		

Interim summary

- It is useful to draw 2-D plots with H1–H2 on one axis and HNR on the other.



Data cleaning for voice quality measures

- Very important and very frequently overlooked
- Things to be cleaned
  - f0
  - vowel formant
  - H1–H2
  - HNR

Data cleaning for voice quality measures

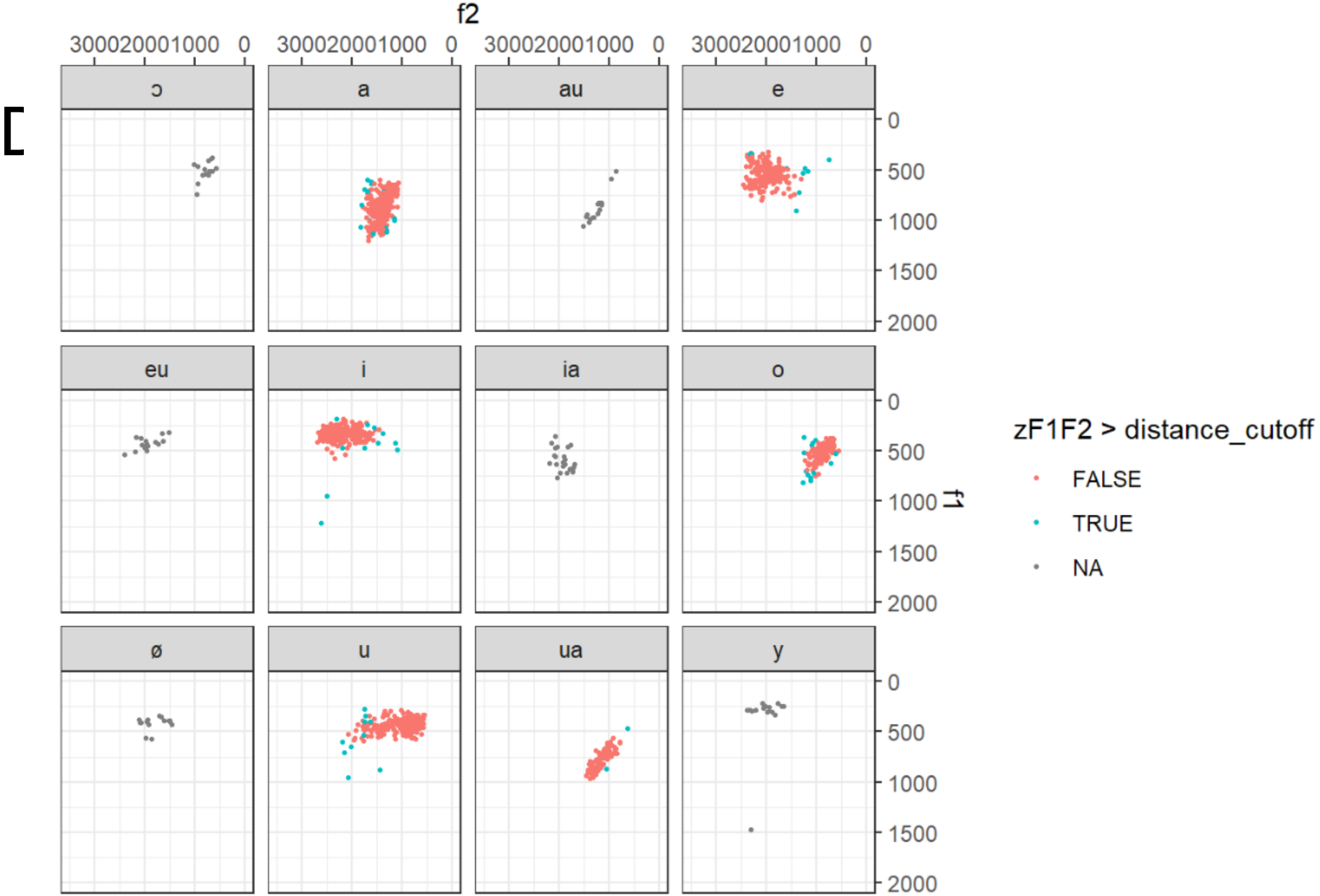
- To clean f0:
- Draw the f0 pitch track out


strF0
1000
750
500
250
0
50
100
150
200
F22May05_LRRH_3

Data cleaning for voice quality measures

- To clean vowel formant:
  - Get the mid-point formant of a vowel
  - Calculate Mahalanobis distance between individual token and the central token of a vowel category

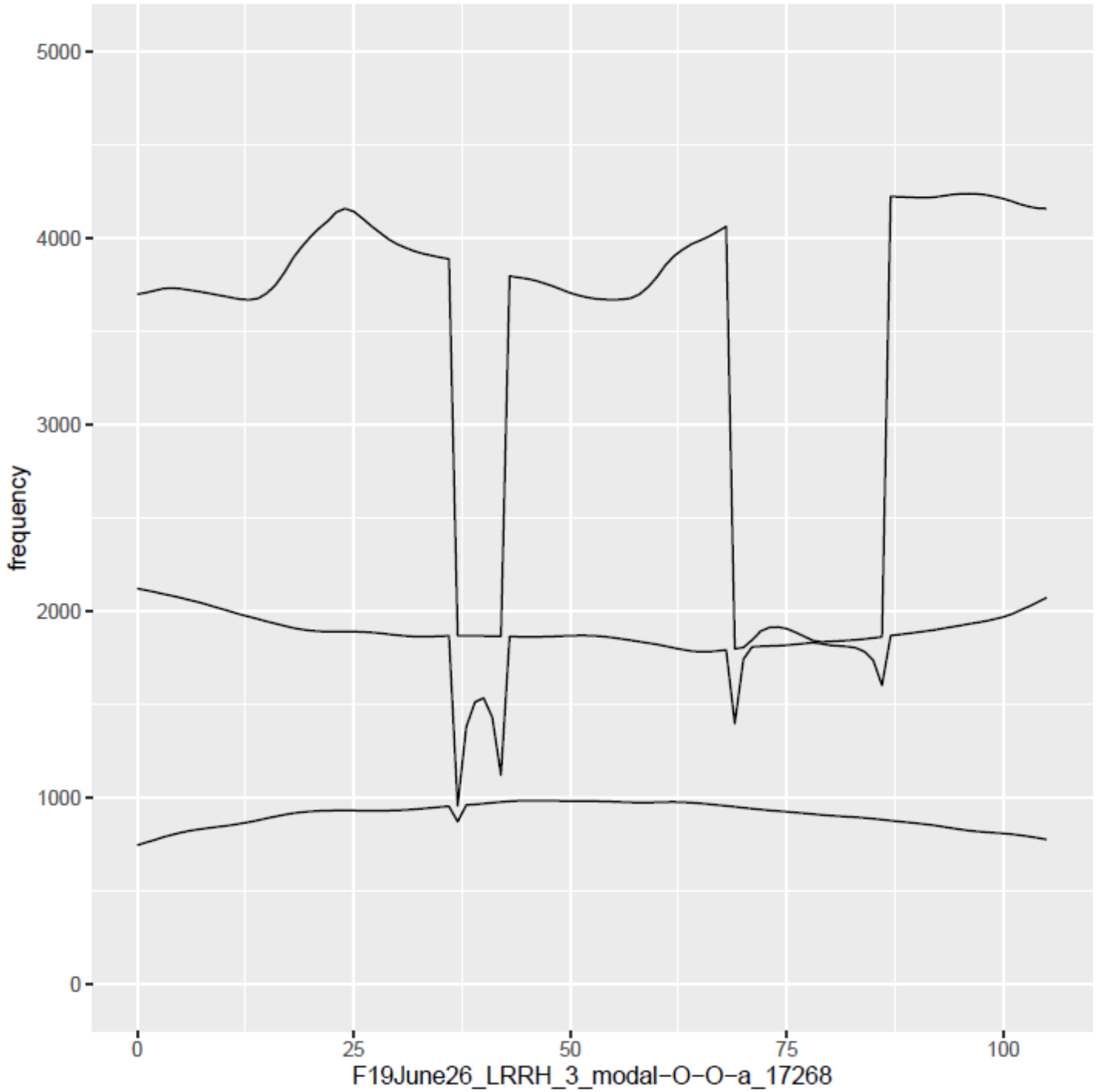
```
means = c(mean(dat$f1), mean(dat$f2))
cov = cov(cbind(dat$f1, dat$f2))
dat$zF1F2 = mahalanobis(cbind(dat$f1, dat$f2), center=means, cov=cov)
```


f2
300020001000 0
ɔ
a
au
e
eu
i
ia
o
ø
u
ua
y
0
500
1000
1500
2000
f1
zF1F2 > distance_cutoff
FALSE
TRUE
NA

Data cleaning for voice quality measures

- To clean vowel formant:
  - Draw out F1, F2, F3 for each token and exclude specific interval manually.

Data cle es


5000
4000
3000
2000
1000
0
frequency
0
25
50
75
100
F19June26_LRRH_3_modal−O−O−a_17268

Data cleaning for voice quality measures

- Suggestion: How to automate the outlier detection process?
  - Looking for jumps in the values?

Data cleaning for voice quality measures

- [u] is problematic
  - Very likely to confuse f0 and F1
  - Might need to do manual checking
  - Or set a threshold for F1
    - If F1 is larger than 1000, discard that token

Data cleaning for voice quality measures

- Clean H1–H2
  - Exclude tokens with f0 tracking errors
  - Exclude tokens with formant tracking errors
- You can further exclude H1–H2 that have standard deviation larger than three.

Data cleaning for voice quality measures

- Harmonic-to-Noise ratio (HNR) and Cepstral Peak Prominence (CPP) are not affected by f0 and formant errors
- You can exclude HNR and CPP values larger than three standard deviations.

Take-home message

- The measure H1–H2, though commonly used, is not always available when you want to measure the degree of vocal fold constriction.
  - If you H1–H2 behaves weirdly, it is very likely because they are not correctly measured.
- CQ is a more direct proxy (though it is still a proxy) for vocal fold constriction.