Friend or Foe?

The 'Critical Role' of Voice Quality in Dungeons and Dragons via Non-Player Characters Voiced by Matt Mercer

Zac Boyd & Míša Hejná

Aarhus Universitet

NWAV50, Stanford 2022

@ZacBoyd_

Dungeons and Dragons

- Dungeons & Dragons (D&D) is a tabletop role-playing game in which players and the Dungeon Master (DM) collaboratively construct a long-form narrative story driven by the actions of the players through gameplay.

- First Edition - 1974

- 5th Edition (5e) Current Release – 2014

- 50+ Million Players Worldwide (Wizards of the Coast (WoTC) 2020)
  - "…We can safely say that D&D is more popular now than it has ever been in it's 47 year history" - Ray Winniner, WoTC D&D Executive Producer, 2020

- 5e has seen a massive resurgence in D&D's popularity – in part thanks to Critical Role

Critical Role

- "Actual Play" DnD Show - Genre of media which broadcasts live Table Top Role Playing Game's

- Cast includes 8 well-established "nerdy-ass voice actors" (Mercer, every single episode of C.R.)

- During their first two D&D campaigns, C.R. produced over 1,000 hours of gameplay over 251 episodes (Currently airing Campaign 3)

- "Historically, C.R.'s Twitch channel has attracted 60,000-75,000 live viewers for each episode… …the total per-episode audience has ranged from 1.2 million to 1.5 million… …the audience has grown more than 23% on Twitch and nearly 50% on YouTube year over year" since it began in 2016 (Spangler 2021)

- *Vox Machina Animated* – Amazon Original series based on Campaign 1

ALL THIS TO SAY: CRITICAL ROLE IS EXTREMELY POPULAR AMONG D&D FANS AND BEYOND

Matthew (Matt) Mercer

All data for the present study comes from voices produced by Matt Mercer.

Specifically focused on C.R. *Campaign 2 – The Mighty Nein.*

- Prolific Voice Actor for Video Games, TV, & Movies – Over 400 IMDB entries

- Dungeon Master (DM) for C.R.

Critical Role – Speech Context

- "High performance" – Coupland (2007)
- The speech setting here is complex and unique
- *Conversationally Performative*
  - Non-scripted performance
  - Guided yet reactionary
  - On the fly switching between profiles which could be akin to numerous *personae*
  - It's a game being played with long term friends... that just so happens to be broadcast to thousands of people

Needless to say, this is not a speaking context in which many people find themselves.

This is a case study where it remains to be seen how generalizable the findings could be.

Core Aim(s)

1) How does Matt Mercer employ aspects of Voice Quality to construct various facets of Non-Player Character (NPC's) identities?

Specifically, what (if any) are the consistencies across characters within this vocal production based on if the character is an **ally** or **enemy** (and by proxy **Good** & **Evil**)?

2) Explore and map out the challenges of a holistic phonatory analysis beyond one/two phonations

The Data

- Non-Player Characters (NPC's) Voiced by Matt Mercer
- C.R. *Campaign 2 – The Mighty Nein*
  - 141 episodes (Each episode is approximately 3-5 Hours in Length)
  - Total of 483.6 hours of Gameplay (CRStats *n.d.*)
    - Full show total time = 556.5 hours
  - Aired 11 January 2018 - 3 June 2021
- 1,144 unique NPC characters introduced during *The Mighty Nein's* Campaign (CRStats *n.d.*)

Time constraints and the fact that I value my sanity prevent a full NPC analysis

The Data

- 19 Characters
  - 12 Male
  - 6 Female
  - 1 Non-/Multi-Gendered 'Hive Mind'
- Range of (D&D) Races
- Mercer's non-character audio comes from episode recaps at the start of the sessions
  - Acts as a 'baseline' for analysis

Name	Total Audio	Alignment	Gender	Creature Type	Race
Artagan (The Traveler)	53:10	Ally	Male	Fey	Archfey
Astrid Becke	10:45	Ally	Female	Humanoid	Human
Avantika	18:28	Ally -> Enemy	Female	Humanoid	Half-Elf
Babenon Dosal (The Gentleman)	53:37	Ally	Male	Humanoid	Genasi
Essek Thelyss	1:53:54	Ally	Male	Humanoid	Drow
Isharnai, the Prism Sage	12:14	Enemy	Female	Fey	Hag
Lucien	45:16	Neutral -> Enemy	Male	Humanoid/ Abberation	Tiefling/ Abberation
Ludinus Da'leth	20:03	Neutral	Male	Humanoid	Elf
Marion Lavorre	35:45	Ally	Female	Humanoid	Tiefling
Obann	10:50	Enemy	Male	Fiend	Cambion
Orly Skiffback	20:46	Ally	Male	Humanoid	Tortle
Pumat Sol	47:55	Ally	Male	Humanoid	Firbolg
The Somnovem	15:23	Enemy	Non-/Multi-Gender	Aberration	Hive-Mind
Trent Ikithon	22:25	Enemy	Male	Humanoid	Human
Uk'otoa	2:26	Enemy	Male	Aberration	Leviathan
Vess DeRogna	36:31	Neutral	Female	Humanoid	Half-Elf
Vilya (Viridian)	41:07	Ally	Female	Humanoid	Half-Elf
Yeza Brenatto	14:40	Ally	Male	Humanoid	Halfling
Yussa Errenis	33:34	Ally	Trans-Male*	Humanoid	Elf
Matt Mercer	04.18				

Auditory Analysis

- Full character audio; listened for:
  - (relative) Modality
  - (relative) Breathiness
  - (relative) Creak
  - (relative) Whisperiness
    - Including growl
    - Including whispery creak
  - Pitch, pitch dynamism, tremor
  - Tempo
  - Constricted supralaryngeal vocal tract
- Combined with visual spectrogram inspection of selected data

Acoustic Analysis

- **At least two** segments per speaker*
  - 1-2 minutes in length
  - Some characters have more than two segments based on evolution of relevant plot points
  - Pseduo-randomly selected from the full speech
    - Selected for 'Emotion' context, with no other considerations

At least one of the following contexts for each character (from here referred to as **'Emotion'**):
- One 'Neutral' – i.e., infomative gameplay related speech
      (Lore building, giving relevant information, setting the party on quests)
- One 'Non-Neutral' – i.e., more emotionally charged
      (Comforting family/friends, post-battle discussions, expressions of anger or nafarious intent)

*(Not possible for Uk'otoa who only has 2m 26s of audio – full audio used)

Acoustic Analysis:
Voice Sauce (Shue 2010; Shue et al. 2011)

- Pseudo-randomly selected (2 minute) segments:
    - Fully transcribed in ELAN (Wittenburg, *et al.* 2006)
    - Aligned via Montreal Forced Aligner (McAuliffe, et al. 2017)
    - Extracted in Voice Sauce
        - F0 (Straight)
        - H1* - H2*; H2* - H4*
        - H* - A1*; H1* - A2*; H1* - A3*
        - Cepstral Peak Prominence - CPP
   - Harmonic to Noise Ratio – HNR

- Results presented today: CPP

Voice Quality as a Stance of Good or *Evil*

- In *all cases*, Enemies (evil) exhibit more whisper and Allies (good) exhibit more breathiness

- Some characters may take stances of general '*evilness*'

  (*e.g.*, Uk'otoa, Isharnai: Essentially signalling 'Here I am. I am Evil.')

  - *Overtly* evil characters employ intense and consistent whispery phonation
    - Extremes showing aryepiglottic trill (i.e. growl)
  - Those with more nuanced personalities employ whispery phonation when the situation is more emotionally charged (*e.g.*, anger/frustration) or conveying character motivation

- More often the stances being taken are (generally) that of *threat* or *safety* (and by extension - *comfort/vulnerability*)

  - *Threat* is realised via whispery phonation; *Safety/comfort/vulnerability* is realised via breathiness
  - In each case, it is either projected towards other or assumed upon the self

# CPP Descriptives
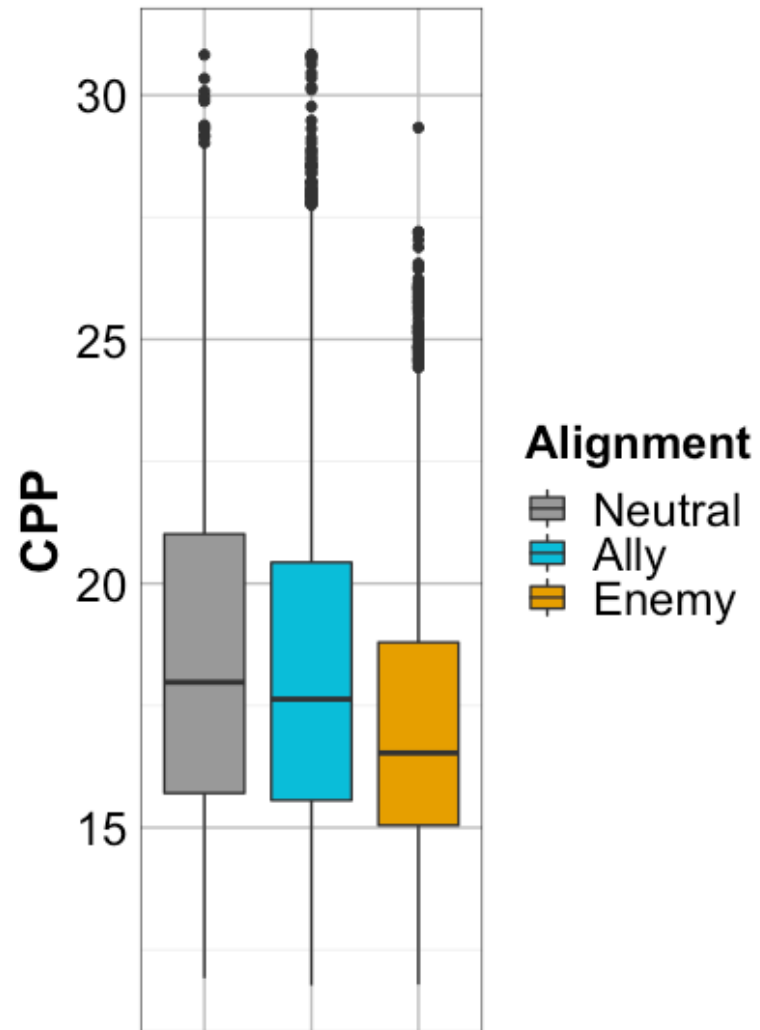
OVERALL: BASED ON ALIGNMENT
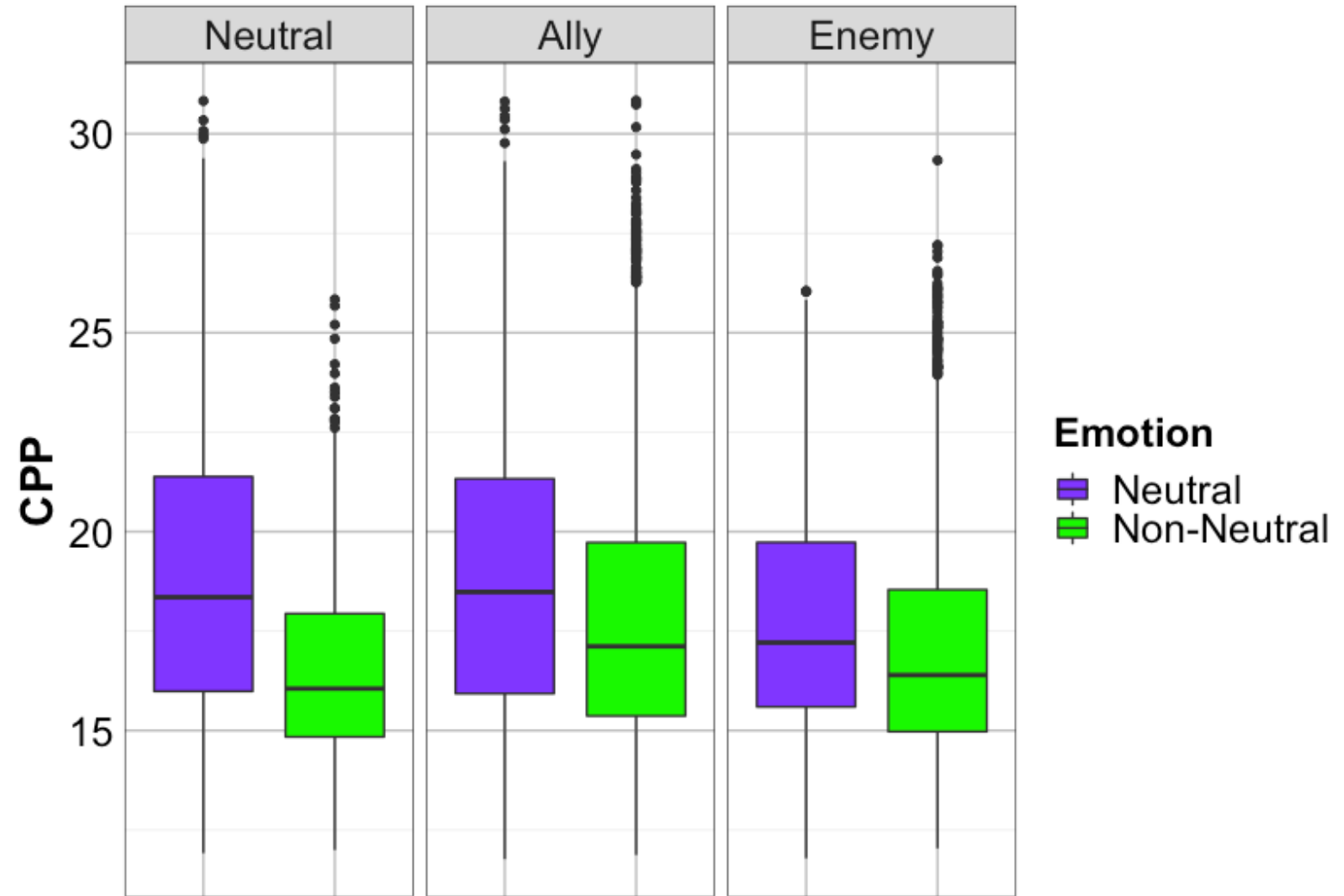
EMOTIONAL CONTEXT BY ALIGNMENT

12

The Modal–Breathy–Whispery Continuum

- Challenges with acoustic analysis whisper/breathy
  - Needs to be done, at least in part, auditorily
  - The measures portray modal – breathy – whispery as a continuum
    - E.g. the lower the CPP, the breathier/whisperier the voice (when contrasted with modal)
    - E.g. the lower the CPP, the whisperier the voice (when contrasted with breathy (and modal))

  - Even so – it's possible to have Breathy & Whispery phonations with the same acoustic values

# Good & Evil: Importance of Auditory Analysis
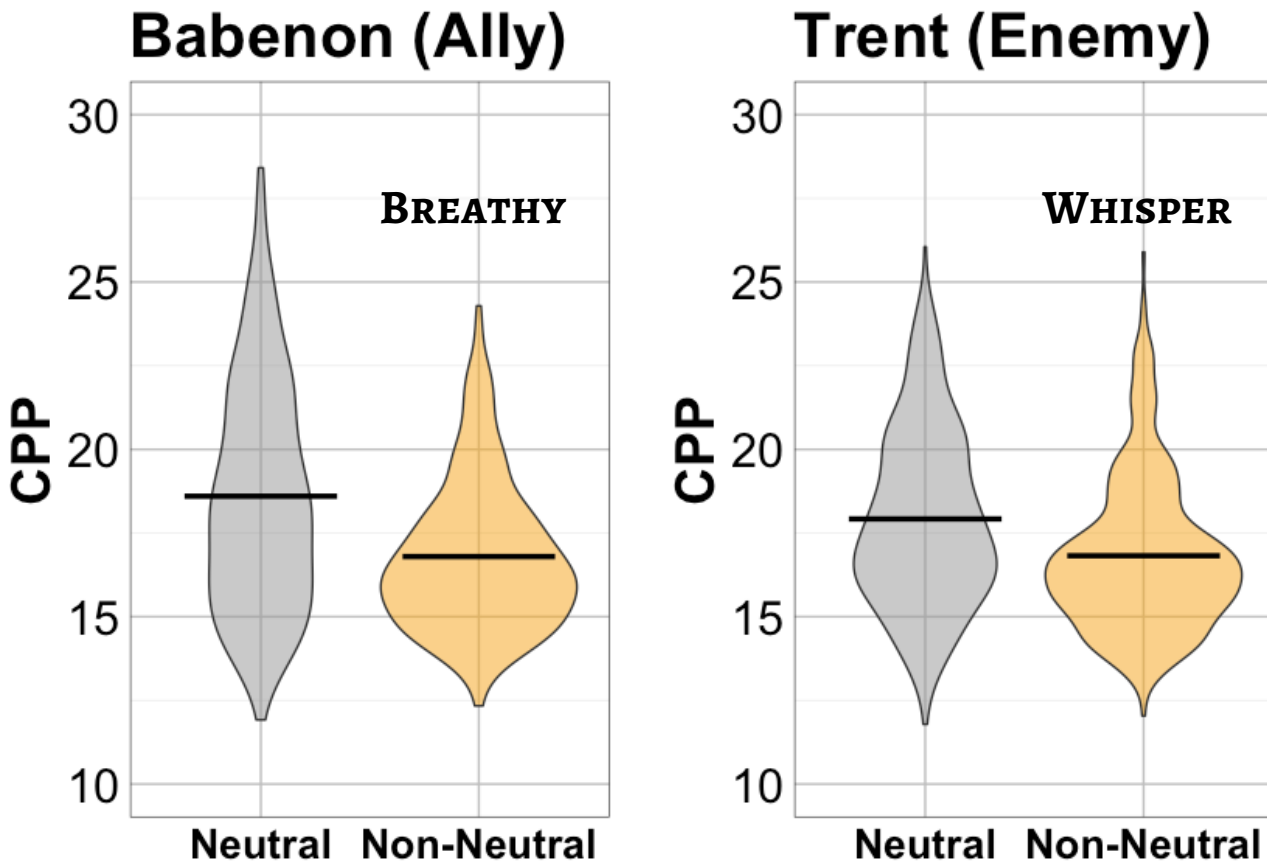
Purely acoustic analyses will have difficulty distinguishing whispery from breathy
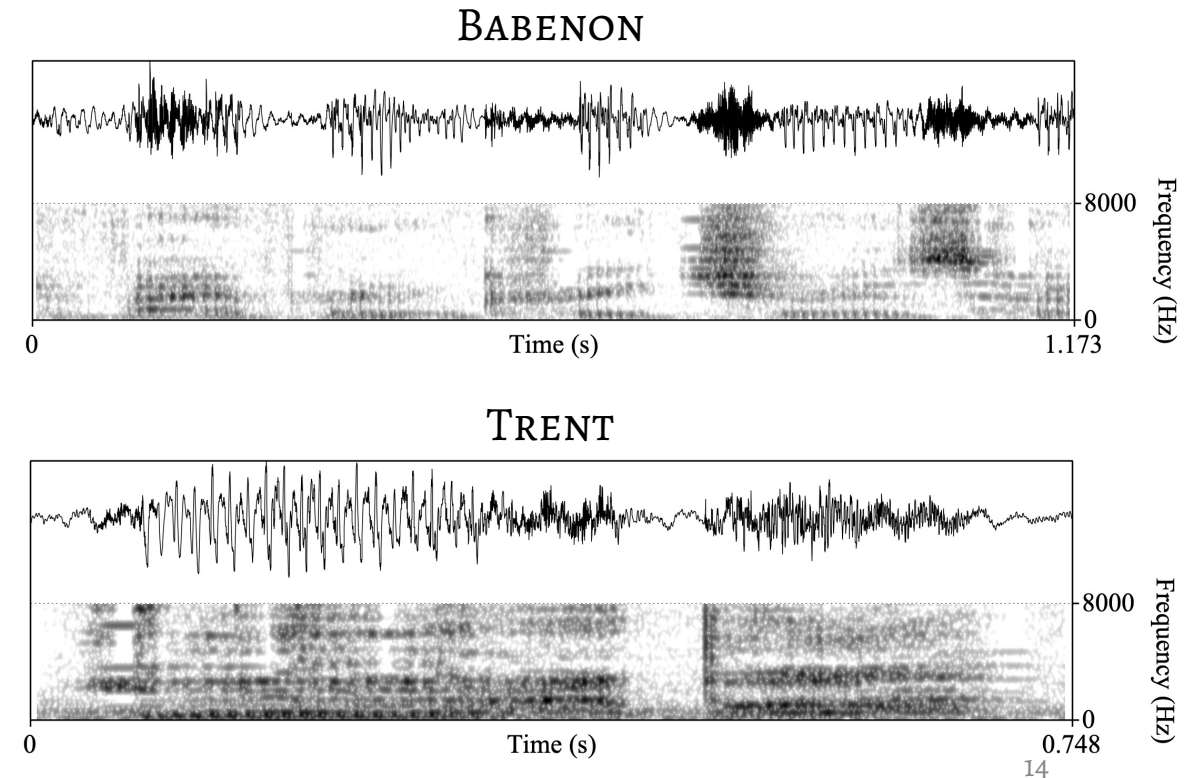
Performing 'Evil'

- Non-human like threats and pathologically affected voices

- Pulmonic Ingressive Airstream
  - Signals non-human-ness

- If you speak with a whispery phonation you are much more likely to be evil and an enemy:
  - The more whispery your whisper is, the more likely you are to be evil/enemy.
    - Esp. when reaching 'growl'

- Whispery – often signalling animal-like (dangerous), non-human, & unhealthy individuals
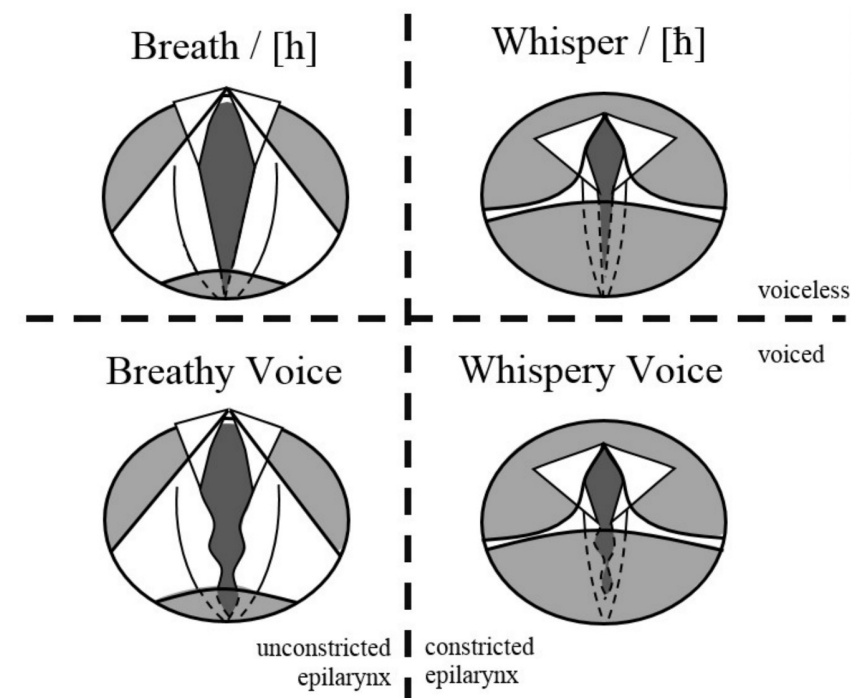
(Kjeldgaard-Christiansen, Hejná, Clasen, & Eaton; *In Press*)

Moisik *et al.* (2019)

15

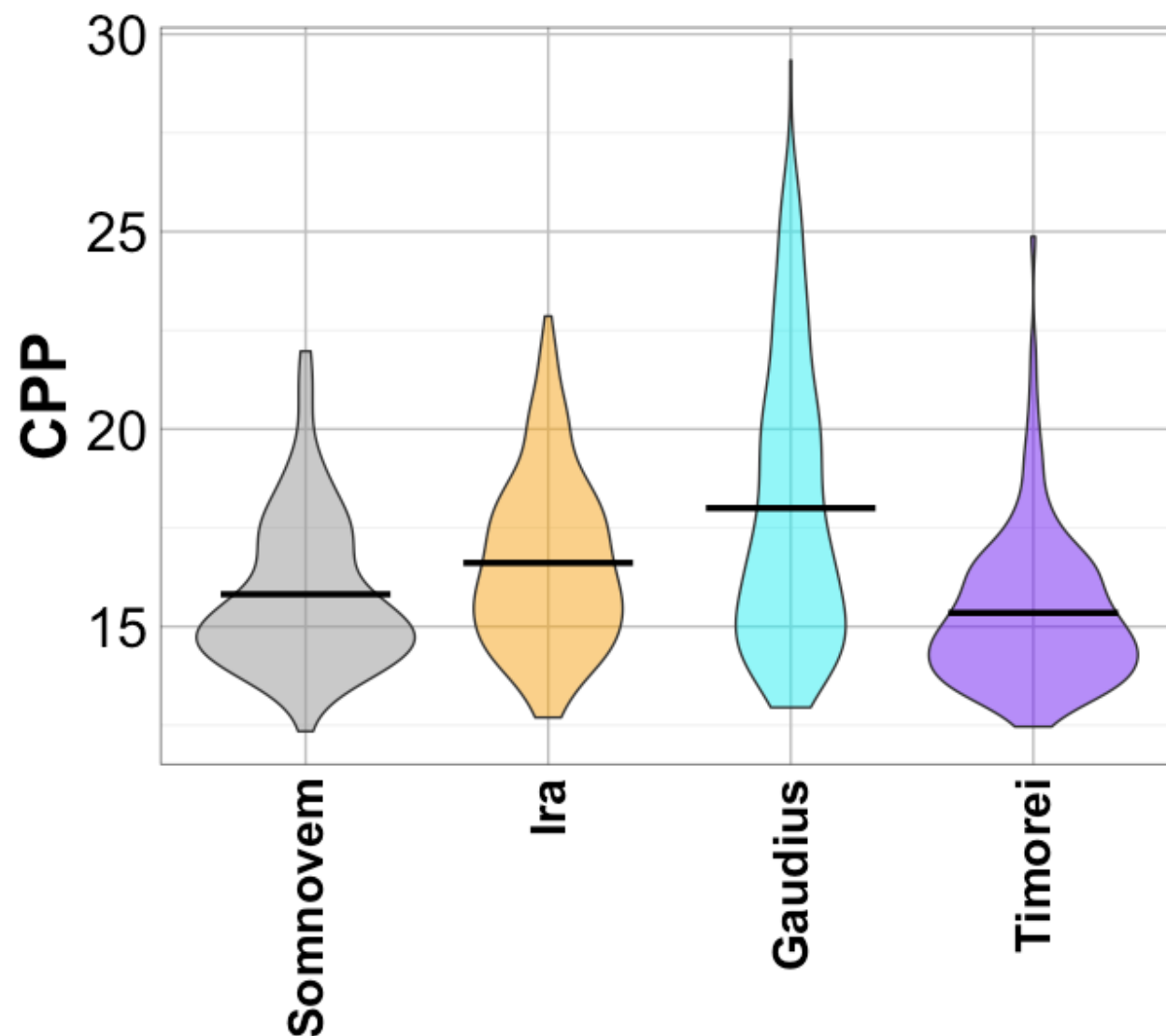# "Damn. Could you, like, be more ominous Orly?"
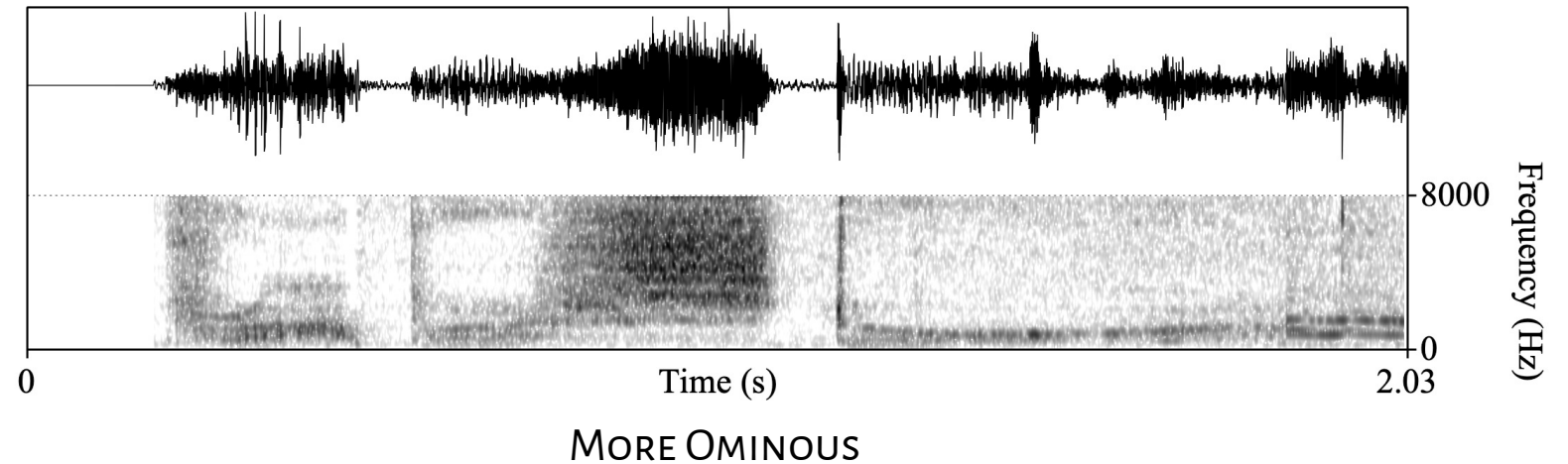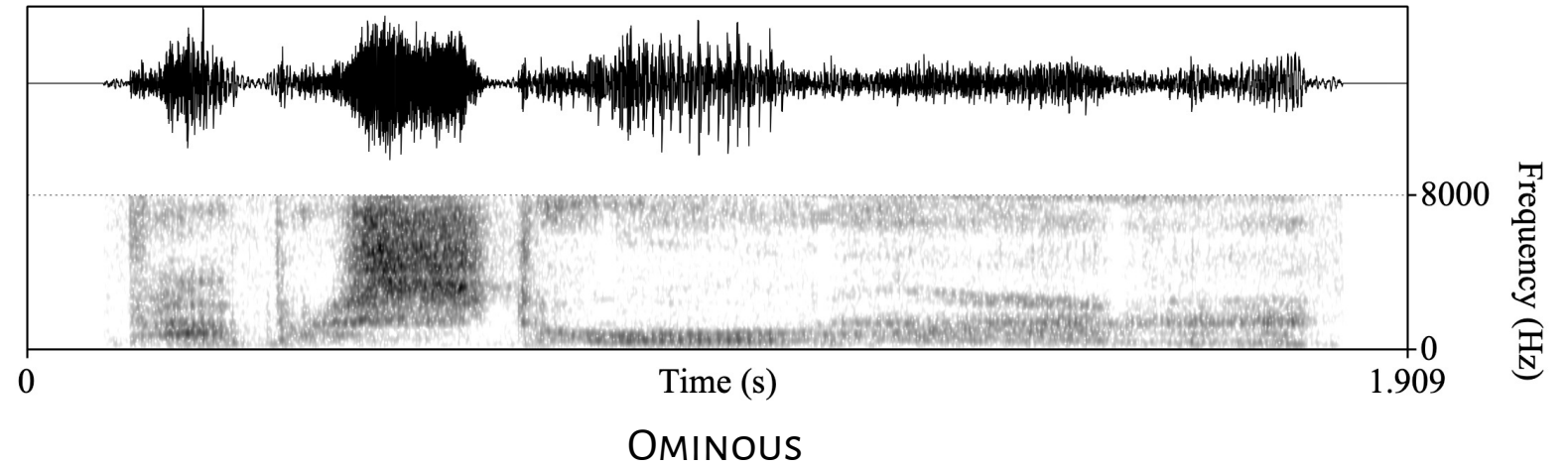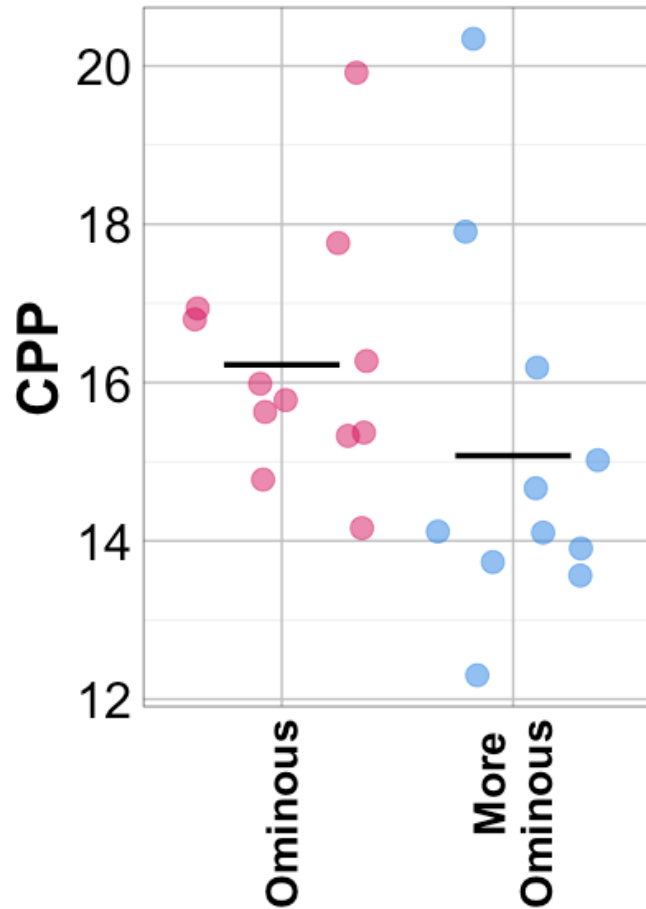
The Hive Mind

A more nuanced approach to evil and phonatory variation

- **Somnovem** – The Encompassing Hive

Single Minds of the Hive:

- **Ira** – '*Wrath*'; harsh intense whisper; consistent 'growl'

- **Gaudius** – '*Love*'; modal with occasional breathy

- **Timorei** – '*Fear*'; very harsh whisper

17

Voice Quality as a Stance of Good or Evil

- "How stances are taken, and which stances are taken, are often habitually repeated by people with similar identities" (Kiesling 2022: 412)
  - Applicable even to this rather unique speaking context
  - The voice qualities used, make sense for the stances being taken
    - (*c.f.*, Teshigawara 2004)
  - We aren't making any revelatory claims or discoveries in terms of stance nor that those phonations are being employed signal good/evil intent.

That said…

The Critical Role of Combined Acoustic and Auditory Analyses

In terms of holistic analyses of phonation:

- There's a trend in (socio)phonetics to heavily rely on acoustic measurements
- The software is not yet at a place to handle holistic accounts of (non?) extreme phonatory variation
- Quasi-periodic nature of 'growl' *may* interfere with the CPP measurements
  - Meaning measurements of speech containing 'growl' may be unreliable

To conduct a holistic voice quality analysis (esp. with extreme variation):

- Qualitative Auditory analyses are *essential* for us to make any (sociolinguistic) claims
  - In our case about performative stances of *Safety* & *Threat* (and by extension 'Good' & 'Evil')

THANK YOU!