Elite Partisan Polarization in Nonverbal Communication

Presentation at EPSA XIII, Glasgow

Mathias Rask — Aarhus University

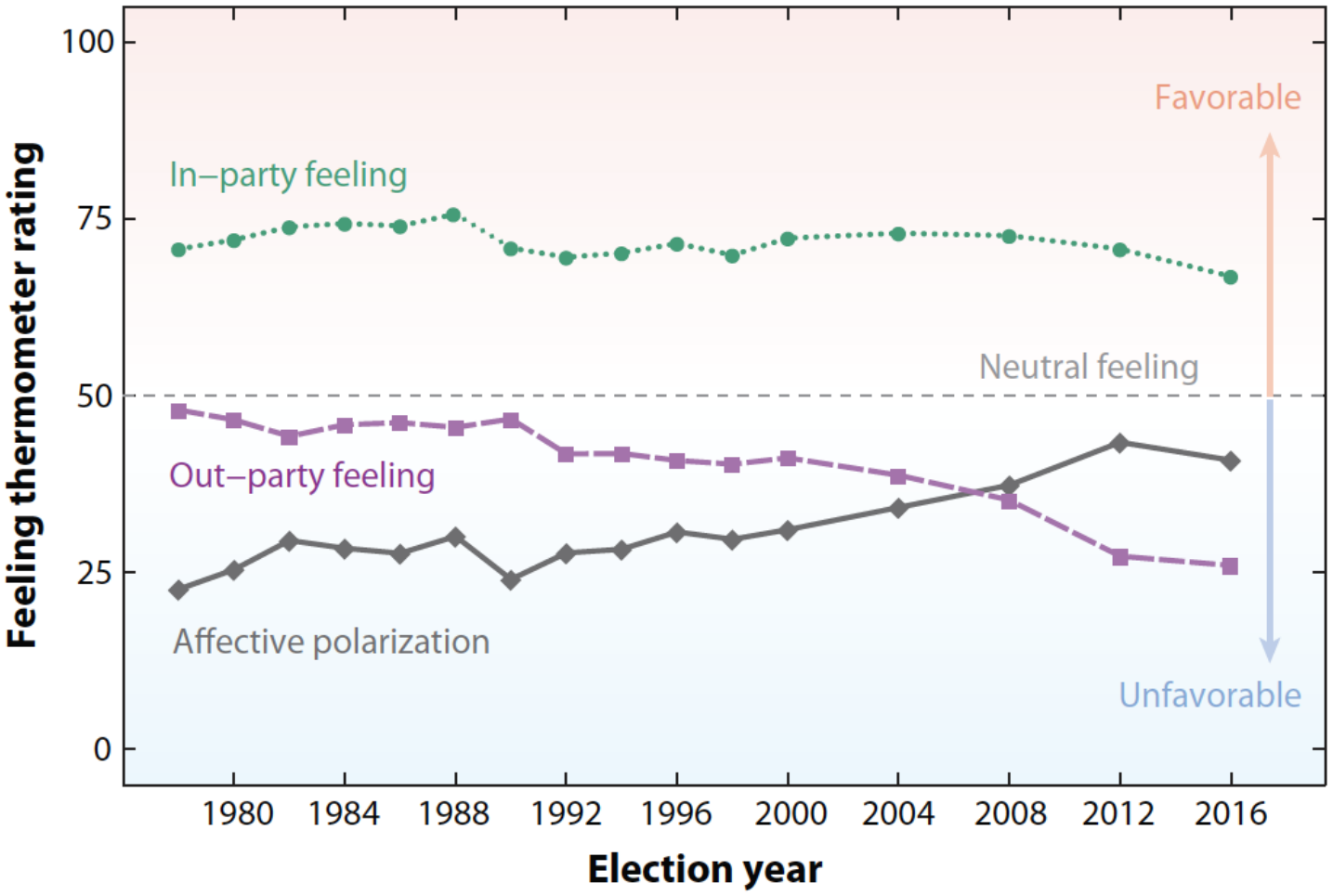
Frederik Hjorth — University of Copenhagen

Motivation

How do elites communicate partisan polarization?

Source: https://www.bbc.com/news/live/uk-politics-63878261

Central to current concerns about affective polarization



Source: Iyengar, S., Lelkes, Y., Levendusky, M., Malhotra, N., & Westwood, S. J. (2019). The origins and consequences of affective polarization in the United States. *Annual review of political science*, 22, 129-146.

Nonverbal signals of elite partisan polarization I

Nonverbal signals of elite partisan polarization I

Source: Andrew Harnik, AP

- we focus on **nonverbal interaction**

Nonverbal signals of elite partisan polarization I

Source: Andrew Harnik, AP

- we focus on **nonverbal interaction**
- essential part of interpersonal communication

Nonverbal signals of elite partisan polarization I

Source: Andrew Harnik, AP

- we focus on **nonverbal interaction**
- essential part of interpersonal communication
- challenge existing literature's focus on verbal content

Nonverbal signals of elite partisan polarization II

Nonverbal signals of elite partisan polarization II

- focus here: vocal style

Nonverbal signals of elite partisan polarization II

- focus here: vocal style
- legislators use vocal style to express agitation at other parties, signaling partisan polarization

Source: https://www.dr.dk/ligetil/se-statsministerens-aabningstale-i-folketinget-paa-100-sekunder & https://www.tv2ostjylland.dk/oestjylland/folketingets-aabning-statsministeren-vil-lave-store-aendringer

Nonverbal signals of elite partisan polarization II

- focus here: vocal style
- legislators use vocal style to express agitation at other parties, signaling partisan polarization
- we are agnostic wrt. intentionality

Source: https://www.dr.dk/ligetil/se-statsministerens-aabningstale-i-folketinget-paa-100-sekunder & https://www.tv2ostjylland.dk/oestjylland/folketingets-aabning-statsministeren-vil-lave-store-aendringer

Hypotheses

Hypotheses

Hypotheses

H_1

Agitation is higher in speech directed at outbloc legislators compared to speech directed at inbloc legislators.

Hypotheses

H_1

Agitation is higher in speech directed at outbloc legislators compared to speech directed at inbloc legislators.

H_2

Agitation is higher in speech directed at legislators with whom they disagree on a bill compared to speech directed at legislators with whom they agree.

Hypotheses

H_1

Agitation is higher in speech directed at outbloc legislators compared to speech directed at inbloc legislators.

H_2

Agitation is higher in speech directed at legislators with whom they disagree on a bill compared to speech directed at legislators with whom they agree. [NOT YET TESTED]

Hypotheses

H_1

Agitation is higher in speech directed at outbloc legislators compared to speech directed at inbloc legislators.

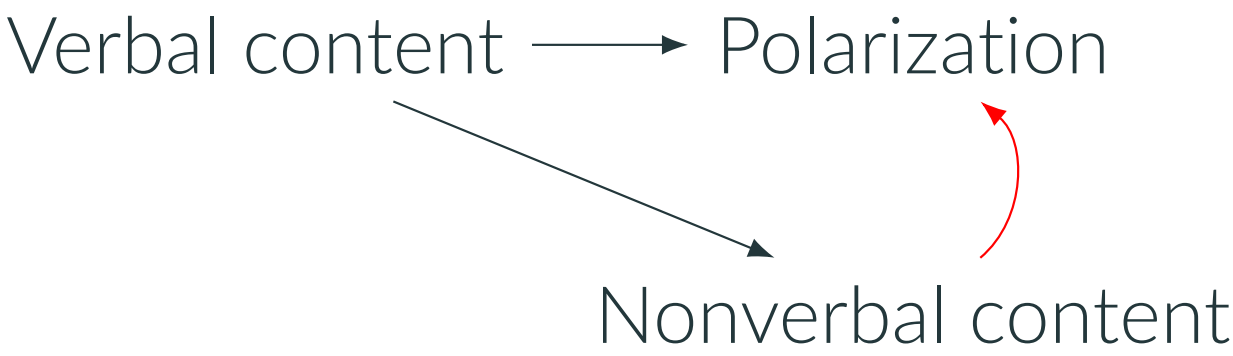
H_2

Agitation is higher in speech directed at legislators with whom they disagree on a bill compared to speech directed at legislators with whom they agree. [NOT YET TESTED]

H_3

Agitation in legislator party-dyads is associated with affective polarization in corresponding party-dyads at the mass level.

Confounding by verbal speech content



Data and Methods

Data Collection: Audio recordings

Data Collection: Audio recordings

- setting: Denmark's parliament *Folketinget*

Data Collection: Audio recordings

- setting: Denmark's parliament *Folketinget*
- 1,302 video recordings btw. Oct. 2010 and Sep. 2022

Data Collection: Audio recordings

- setting: Denmark's parliament *Folketinget*
- 1,302 video recordings btw. Oct. 2010 and Sep. 2022
- avg. length ≈ 5 hours

Data Collection: Audio recordings

- setting: Denmark's parliament *Folketinget*
- 1,302 video recordings btw. Oct. 2010 and Sep. 2022
- avg. length ≈ 5 hours
- remove recordings with only chairperson speaking

Data Collection: Audio recordings

- setting: Denmark's parliament *Folketinget*
- 1,302 video recordings btw. Oct. 2010 and Sep. 2022
- avg. length ≈ 5 hours
- remove recordings with only chairperson speaking
- extract audio channel from video

Data Collection: Audio recordings

- setting: Denmark's parliament *Folketinget*
- 1,302 video recordings btw. Oct. 2010 and Sep. 2022
- avg. length $\approx$ 5 hours
- remove recordings with only chairperson speaking
- extract audio channel from video
- align audio to speech text using alignment method developed in separate working paper [6]

Data Collection: Audio recordings

- setting: Denmark's parliament *Folketinget*
- 1,302 video recordings btw. Oct. 2010 and Sep. 2022
- avg. length $\approx$ 5 hours
- remove recordings with only chairperson speaking
- extract audio channel from video
- align audio to speech text using alignment method developed in separate working paper [6]

Data Collection: Audio recordings

- setting: Denmark's parliament *Folketinget*
- 1,302 video recordings btw. Oct. 2010 and Sep. 2022
- avg. length ≈ 5 hours
- remove recordings with only chairperson speaking
- extract audio channel from video
- align audio to speech text using alignment method developed in separate working paper [6]

→ linked audio and text for ≈ 209,000 speeches

Data Collection: Audio recordings

- setting: Denmark's parliament *Folketinget*
- 1,302 video recordings btw. Oct. 2010 and Sep. 2022
- avg. length ≈ 5 hours
- remove recordings with only chairperson speaking
- extract audio channel from video
- align audio to speech text using alignment method developed in separate working paper [6]

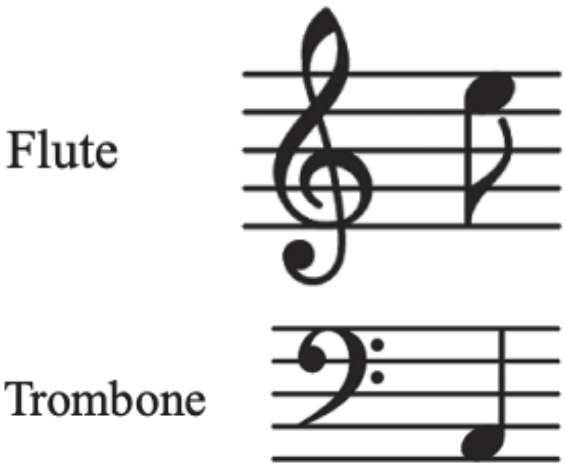
→ linked audio and text for ≈ 209,000 speeches

▸ Additional detail on alignment

Measurement I: Measuring agitation

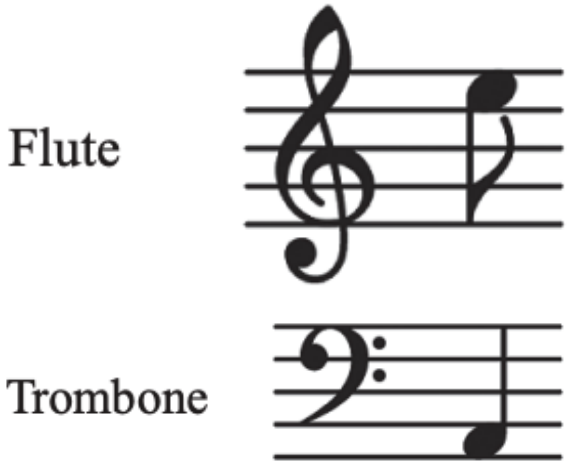
Measurement I: Measuring agitation

- use standardized vocal pitch as proxy



High vs. low pitch illustration from Knox & Lucas [5]

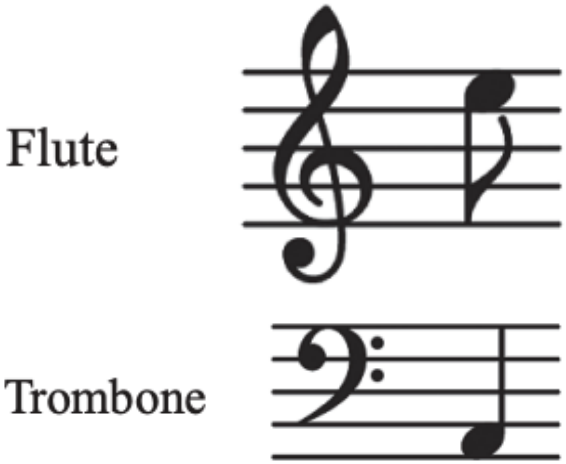
Measurement I: Measuring agitation



High vs. low pitch illustration from Knox & Lucas [5]

- use standardized vocal pitch as proxy
- context-dependent meaning: pitch used in earlier work to measure issue commitment [4] and judicial intent [3]

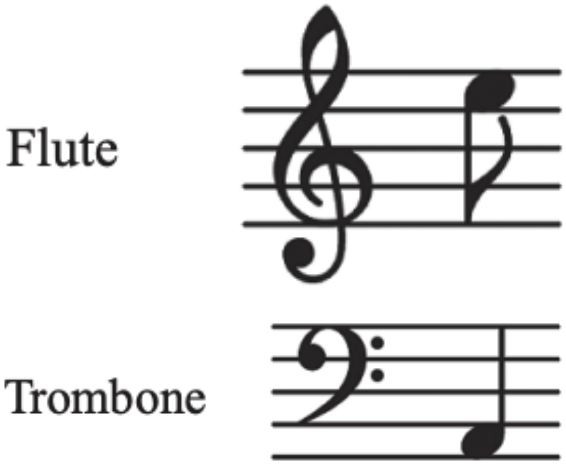
Measurement I: Measuring agitation



High vs. low pitch illustration from Knox & Lucas [5]

- use standardized vocal pitch as proxy
- context-dependent meaning: pitch used in earlier work to measure issue commitment [4] and judicial intent [3]
- standardize pitch within each speaker to remove speaker heterogeneity (e.g. gender)

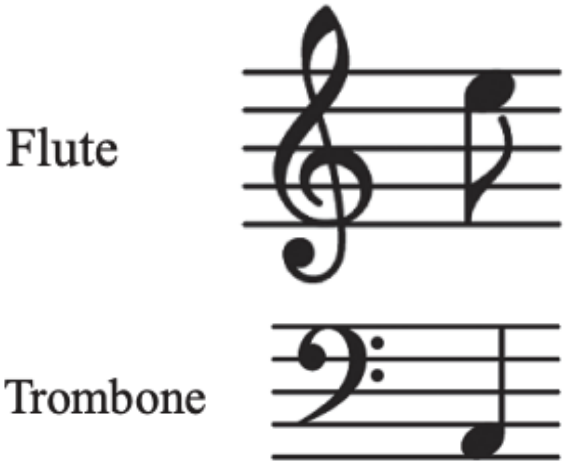
Measurement I: Measuring agitation



High vs. low pitch illustration from Knox & Lucas [5]

- use standardized vocal pitch as proxy
- context-dependent meaning: pitch used in earlier work to measure issue commitment [4] and judicial intent [3]
- standardize pitch within each speaker to remove speaker heterogeneity (e.g. gender)
- $\rightsquigarrow$ implicit control for all fixed legislator-level characteristics

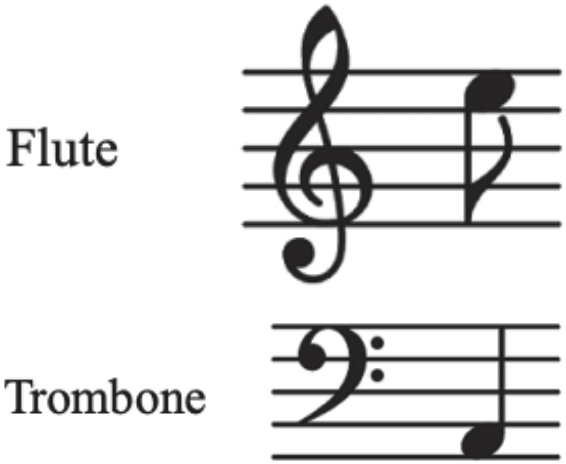
Measurement I: Measuring agitation



High vs. low pitch illustration from Knox & Lucas [5]

- use standardized vocal pitch as proxy
- context-dependent meaning: pitch used in earlier work to measure issue commitment [4] and judicial intent [3]
- standardize pitch within each speaker to remove speaker heterogeneity (e.g. gender)
- $\rightsquigarrow$ implicit control for all fixed legislator-level characteristics
- **validation**: higher pitch correlates with (a) manual coding of agitated speech and (b) negativity in verbal content

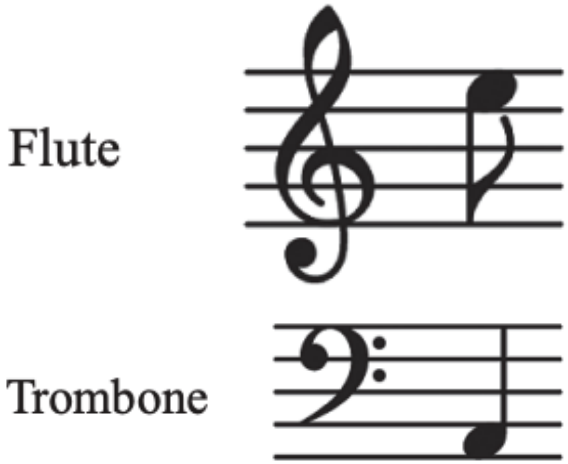
Measurement I: Measuring agitation



High vs. low pitch illustration from Knox & Lucas [5]

- use standardized vocal pitch as proxy
- context-dependent meaning: pitch used in earlier work to measure issue commitment [4] and judicial intent [3]
- standardize pitch within each speaker to remove speaker heterogeneity (e.g. gender)
- $\rightsquigarrow$ implicit control for all fixed legislator-level characteristics
- **validation**: higher pitch correlates with (a) manual coding of agitated speech and (b) negativity in verbal content

Measurement I: Measuring agitation

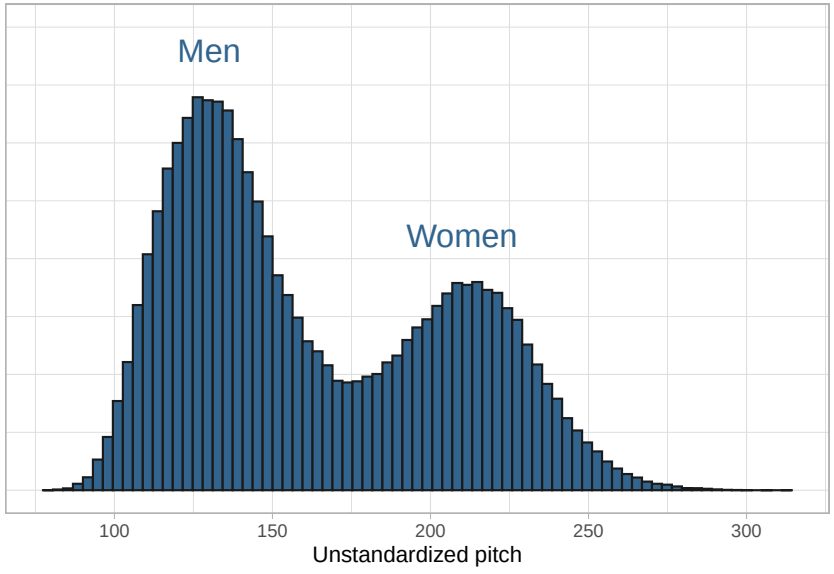


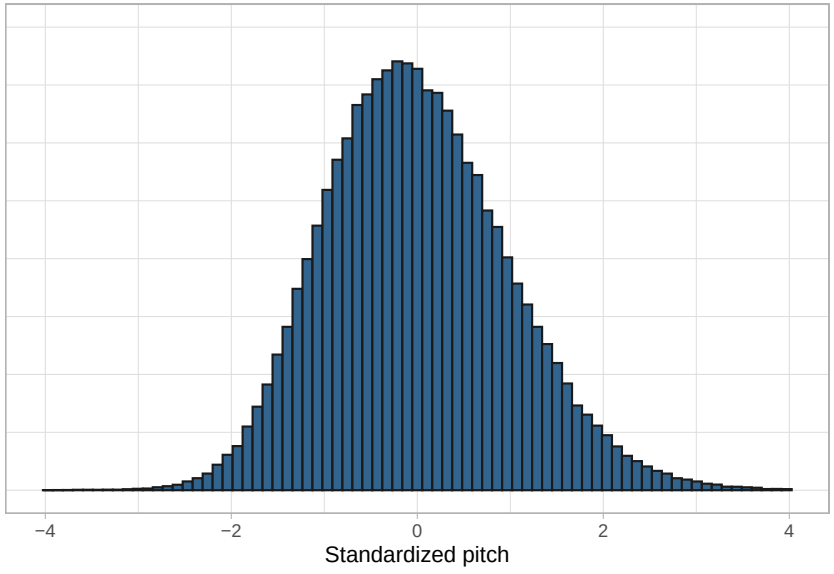
High vs. low pitch illustration from Knox & Lucas [5]

- use standardized vocal pitch as proxy
- context-dependent meaning: pitch used in earlier work to measure issue commitment [4] and judicial intent [3]
- standardize pitch within each speaker to remove speaker heterogeneity (e.g. gender)
- $\rightsquigarrow$ implicit control for all fixed legislator-level characteristics
- **validation**: higher pitch correlates with (a) manual coding of agitated speech and (b) negativity in verbal content

▸ Additional detail on measurement ▸ Validation ▸ Low pitch example ▸ High pitch example

Distributions of unstandardized and standardized pitch





Measurement II: Identifying party dyads

Measurement II: Identifying party dyads

- find speeches where speakers mentions (i) one or more speakers from one other party, and (ii) no speakers from any other party

Measurement II: Identifying party dyads

- find speeches where speakers mentions (i) one or more speakers from one other party, and (ii) no speakers from any other party
- in these, define speaker party $\Rightarrow$ target party as party dyad

Measurement II: Identifying party dyads

- find speeches where speakers mentions (i) one or more speakers from one other party, and (ii) no speakers from any other party
- in these, define speaker party $\Rightarrow$ target party as party dyad
- identifies party dyad in $\approx$37 pct. of speeches

Measurement II: Identifying party dyads

- find speeches where speakers mentions (i) one or more speakers from one other party, and (ii) no speakers from any other party
- in these, define speaker party $\Rightarrow$ target party as party dyad
- identifies party dyad in $\approx$37 pct. of speeches
- in H_3, measure dyad-level affective polarization by aggregating party sympathy scale in survey data by respondent party affiliation

Measurement III: Controls for verbal speech content

Measurement III: Controls for verbal speech content

- Sentiment, using dictionary-based sentiment model `Sentida`

Measurement III: Controls for verbal speech content

- Sentiment, using dictionary-based sentiment model `Sentida`
- Emotionality, using word embeddings approach from Gennaro & Ash (2022)

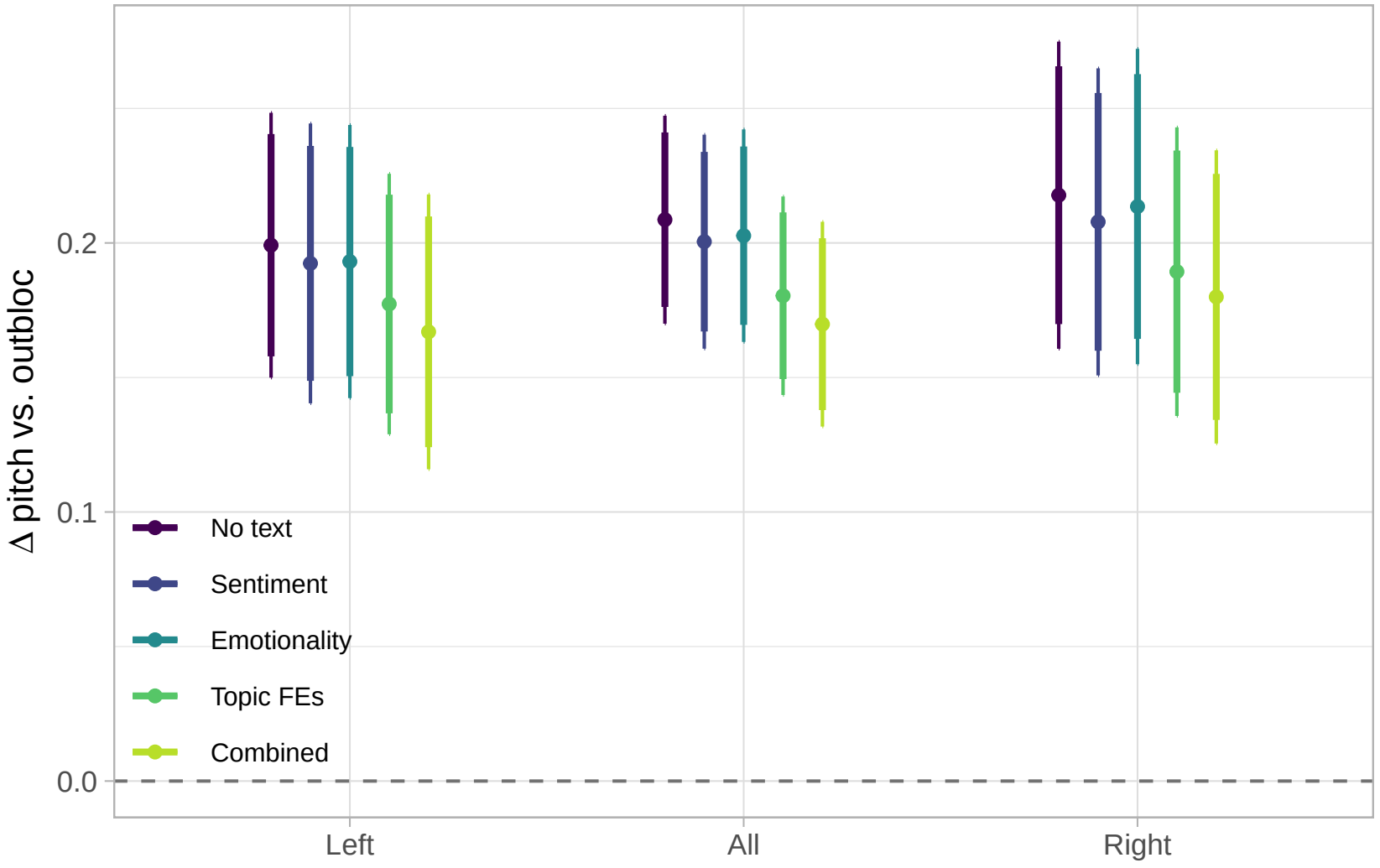
Measurement III: Controls for verbal speech content

- Sentiment, using dictionary-based sentiment model `Sentida`
- Emotionality, using word embeddings approach from Gennaro & Ash (2022)
- Topic fixed effects, using Structural Topic Model w. $k = 40$

Results

H_1: Agitation vs. outbloc targets

H_1: Agitation vs. outbloc targets

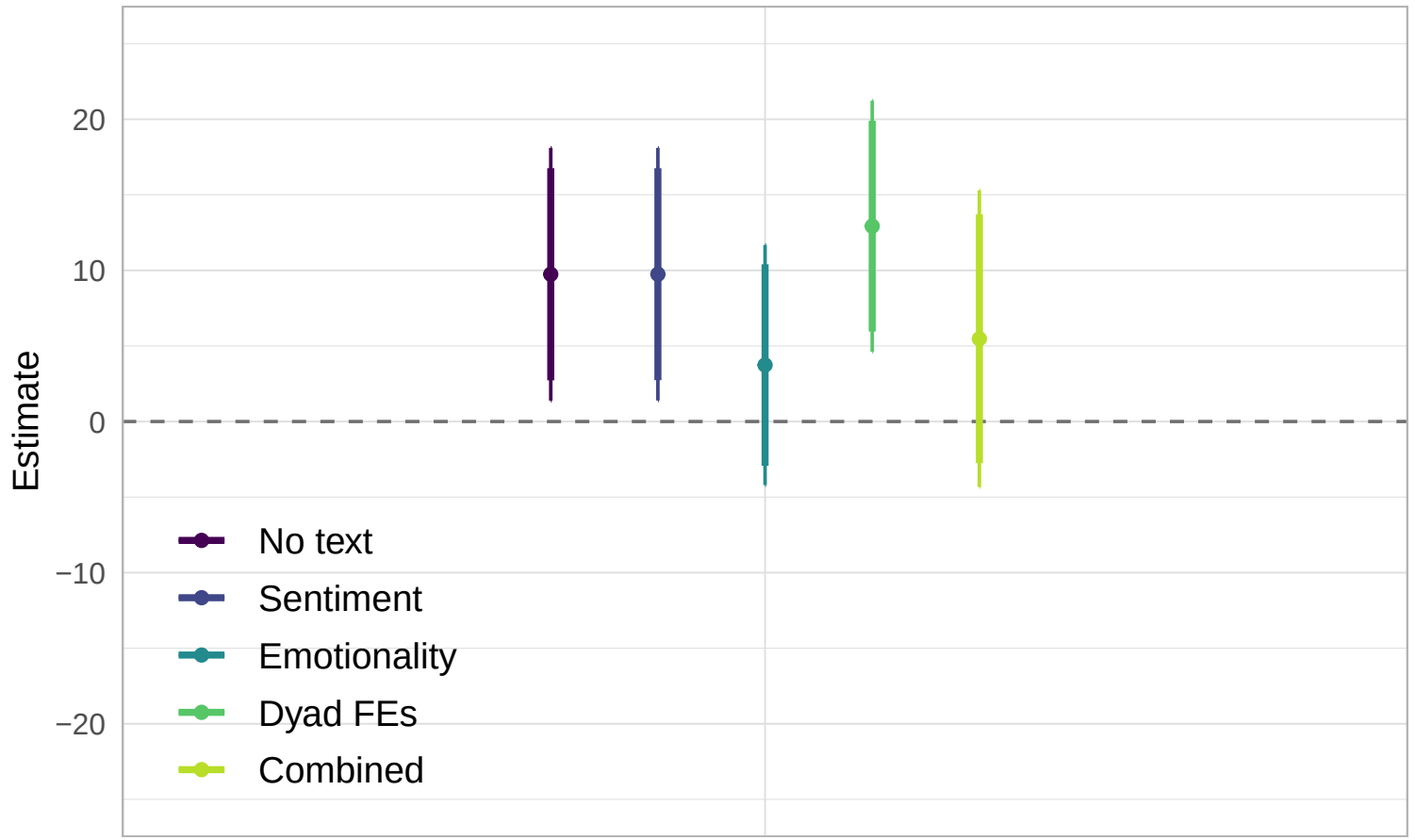


H_2: Agitation and legislative votes

H_2: Agitation and legislative votes

[NOT YET TESTED]

H_3: Agitation and mass affective polarization



Conclusion and Implications

Conclusion

Conclusion

- elite partisan polarization is reflected in legislators' vocal style

Conclusion

- elite partisan polarization is reflected in legislators' vocal style
- vocal style variation across party dyads is reflected in dyad-level affective polarization

Conclusion

- elite partisan polarization is reflected in legislators' vocal style
- vocal style variation across party dyads is reflected in dyad-level affective polarization
- both (largely) hold when accounting for verbal content of communication

Conclusion

- elite partisan polarization is reflected in legislators' vocal style
- vocal style variation across party dyads is reflected in dyad-level affective polarization
- both (largely) hold when accounting for verbal content of communication
- $\Longrightarrow$ nonverbal communication accounts for a distinct dimension of elite communication of partisan polarization

Conclusion

- elite partisan polarization is reflected in legislators' vocal style
- vocal style variation across party dyads is reflected in dyad-level affective polarization
- both (largely) hold when accounting for verbal content of communication
- $\Longrightarrow$ nonverbal communication accounts for a distinct dimension of elite communication of partisan polarization
- caveats:

Conclusion

- elite partisan polarization is reflected in legislators' vocal style
- vocal style variation across party dyads is reflected in dyad-level affective polarization
- both (largely) hold when accounting for verbal content of communication
- $\Longrightarrow$ nonverbal communication accounts for a distinct dimension of elite communication of partisan polarization
- caveats:
  - crude measure of vocal style

Conclusion

- elite partisan polarization is reflected in legislators' vocal style
- vocal style variation across party dyads is reflected in dyad-level affective polarization
- both (largely) hold when accounting for verbal content of communication
- $\Longrightarrow$ nonverbal communication accounts for a distinct dimension of elite communication of partisan polarization
- caveats:
  - crude measure of vocal style
  - observational, cross-sectional design

Implications

Implications

- $\rightsquigarrow$ any assessment of partisan polarization among elites based on verbal content alone is incomplete

Implications

- $\rightsquigarrow$ any assessment of partisan polarization among elites based on verbal content alone is incomplete
- $\rightsquigarrow$ efforts to encourage civil, bipartisan, conciliatory behavior should consider both verbal and nonverbal dimensions of communication

Implications

- $\rightsquigarrow$ any assessment of partisan polarization among elites based on verbal content alone is incomplete
- $\rightsquigarrow$ efforts to encourage civil, bipartisan, conciliatory behavior should consider both verbal and nonverbal dimensions of communication
- $\rightsquigarrow$ individual differences in vocal style can condition political representation

Thanks for your attention!

Elite Partisan Polarization in Nonverbal Communication

Mathias Rask & Frederik Hjorth

References i


H. Bredin and A. Laurent.
End-to-end speaker segmentation for overlap-aware resegmentation.
arXiv preprint arXiv:2104.04045, 2021.


H. Bredin, R. Yin, J. M. Coria, G. Gelly, P. Korshunov, M. Lavechin, D. Fustes, H. Titeux, W. Bouaziz, and M.-P. Gill.
pyannote.audio: neural building blocks for speaker diarization.
In *ICASSP 2020, IEEE International Conference on Acoustics, Speech, and Signal Processing*, Barcelona, Spain, May 2020.


B. J. Dietrich, R. D. Enos, and M. Sen.
Emotional arousal predicts voting on the US supreme court.
Political Analysis, 27(2):237–243, 2019.

References ii


B. J. Dietrich, M. Hayes, and D. Z. O'brien.
Pitch perfect: Vocal pitch and the emotional intensity of congressional speech.
American Political Science Review, 113(4):941–962, 2019.


D. Knox and C. Lucas.
A dynamic model of speech for the social sciences.
American Political Science Review, 115(2):649–666, 2021.


M. Rask.
PolAnnotate: Matching Audio to Transcripts.
Working Paper, 2023.


D. Snyder, D. Garcia-Romero, G. Sell, D. Povey, and S. Khudanpur.
X-vectors: Robust dnn embeddings for speaker recognition.
In *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pages 5329–5333. IEEE, 2018.

Tidying audio: Annotation I

New method and Python software developed in separate working paper (Rask, 2023): `polannote`

- Automated annotation of audio recordings using weak-supervision (i.e. no prior human-annotated data)

Tidying audio: Annotation I

New method and Python software developed in separate working paper (Rask, 2023): `polannote`

- Automated annotation of audio recordings using weak-supervision (i.e. no prior human-annotated data)
- A single annotation: A tuple of start and stop timestamps, speaker name, and the text of speech

Tidying audio: Annotation I

New method and Python software developed in separate working paper (Rask, 2023): `polannote`

- Automated annotation of audio recordings using weak-supervision (i.e. no prior human-annotated data)
- A single annotation: A tuple of start and stop timestamps, speaker name, and the text of speech
- Input: One recording, one transcript

Tidying audio: Annotation I

New method and Python software developed in separate working paper (Rask, 2023): `polannote`

- Automated annotation of audio recordings using weak-supervision (i.e. no prior human-annotated data)
- A single annotation: A tuple of start and stop timestamps, speaker name, and the text of speech
- Input: One recording, one transcript
- Output: A list of annotations (for each recording)

Tidying audio: Annotation II

Back to presentation

Data preprocessing and requirements

- Audio: Divide recording into K batches

$$K = \left\lceil \frac{\text{number of samples/sampling rate}}{60 \times \text{batch duration}} \right\rceil$$

  - Done to lower computational cost
- Transcript: Ordered collection of speeches contained in the audio (e.g. ParlSpeech V2). Three needs:
  - Ordering: $i < i + 1$
  - Speaker names
  - Text

$\longrightarrow$ Method requires the existence of a corresponding transcript

Tidying audio: Annotation III

Method: Two steps

- Step 1: Speaker diarization
- Step 2: Speaker and speech recognition

Tidying audio: Annotation IV

Step 1: Speaker diarization

- Segments a recording into N individual speeches using unsupervised learning
- State-of-the-art software: `pyannote.audio` [2, 1] using neural network building blocks
- Output: A list of speech segments each with corresponding sets of timestamps and generic speaker labels (e.g. A, B, C, etc.).
    - Due to batching, speaker A in batch k is generally not the same as speaker A in batch $k + 1$

Tidying audio: Annotation V

Step 2: Speaker and speech recognition

- Due to the transcript, we know the target speaker and text *ex ante*
  $\longrightarrow$ Weak supervision!
- Construct supervisory signals from each segment from step 1
  - Text signal: Automatic Speech Recognition (ASR) on each segment to generate hypothesis text
  - Audio signal: Generate 512-dimensional embeddings (last layer of a x-vector TDNN-based architecture [7]) for each segment.
    Embeddings encode the speaker identity.
- Match supervisory signals to targets using fuzzy linkages
  - Cosine similarity for text signal – text target (match on words) and for audio signal – audio target (match on speaker)
  - Assign to target if above threshold(s)

Tidying audio: Annotation VI

Analyze validity of method by manually annotating a recording with ground-truth timestamps.

Evaluation metric: Diarization Error Metric (DER)

$$\text{DER} = \frac{\text{false alarm} + \text{missed detection} + \text{confusion}}{\text{total}}$$

- Total: Total duration of ground-truth speaker time
- False alarm: total duration of speech not within the ground-truth timestamps
- Missed detection: total duration of ground-truth speech falsely assigned as non-speech (non-speech: timestamps that fall outside ground-truth)
- Confusion: total duration of speech assigned to a wrong speaker.

The false alarm and missed detection capture the quality of the diarization in step 1 and confusion capture the quality of the speaker

Tidying audio: Annotation VII

Ground-truth: Recording from Danish parliament on December 6 2012 with 478 speeches

- DER: 1.6%
- Using official timestamps (provided in metadata): 17.6%

Measurement II: Measuring indignation

Back to presentation

Standardized vocal pitch as a proxy for indignation:

- estimate pitch for all speeches using open-source software `communication` (R package by [5])
  - estimates are computed on 25 ms windows with 12.5 ms overlap (= 800 estimates on 10 seconds of audio)
  - pitch is tracked by two algorithms: We consider a window as valid if both algorithms return estimates > 0
  - compute mean of valid pitch estimates for each speech to obtain a speech-level measure
- standardize pitch within each speaker to remove speaker heterogeneity (e.g. gender)

Validation I: Manual annotation

Back to presentation

Validation I: Manual annotation

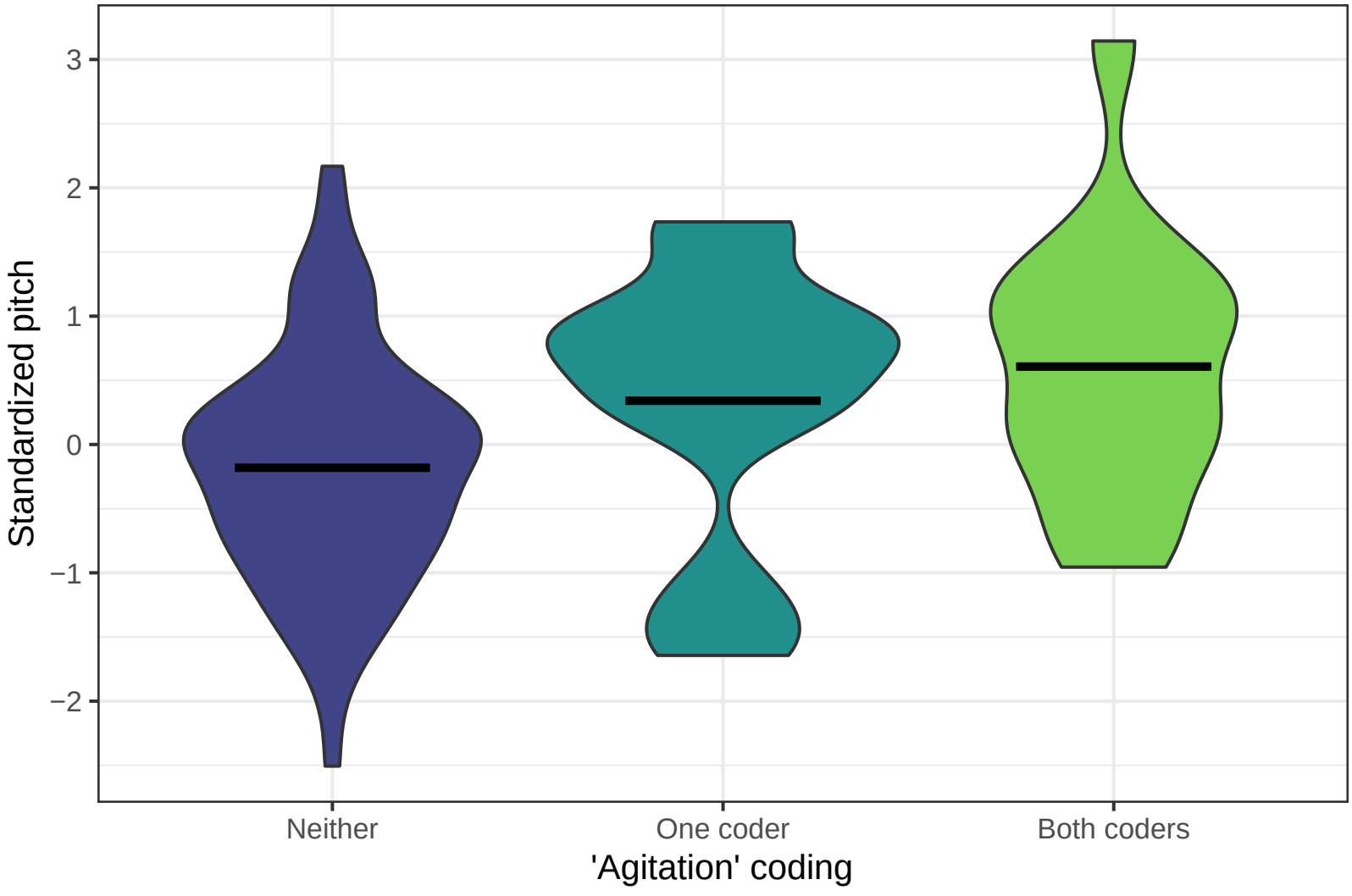
Back to presentation

Manual annotation of 100 speeches for presence of agitation:

Validation I: Manual annotation

Back to presentation

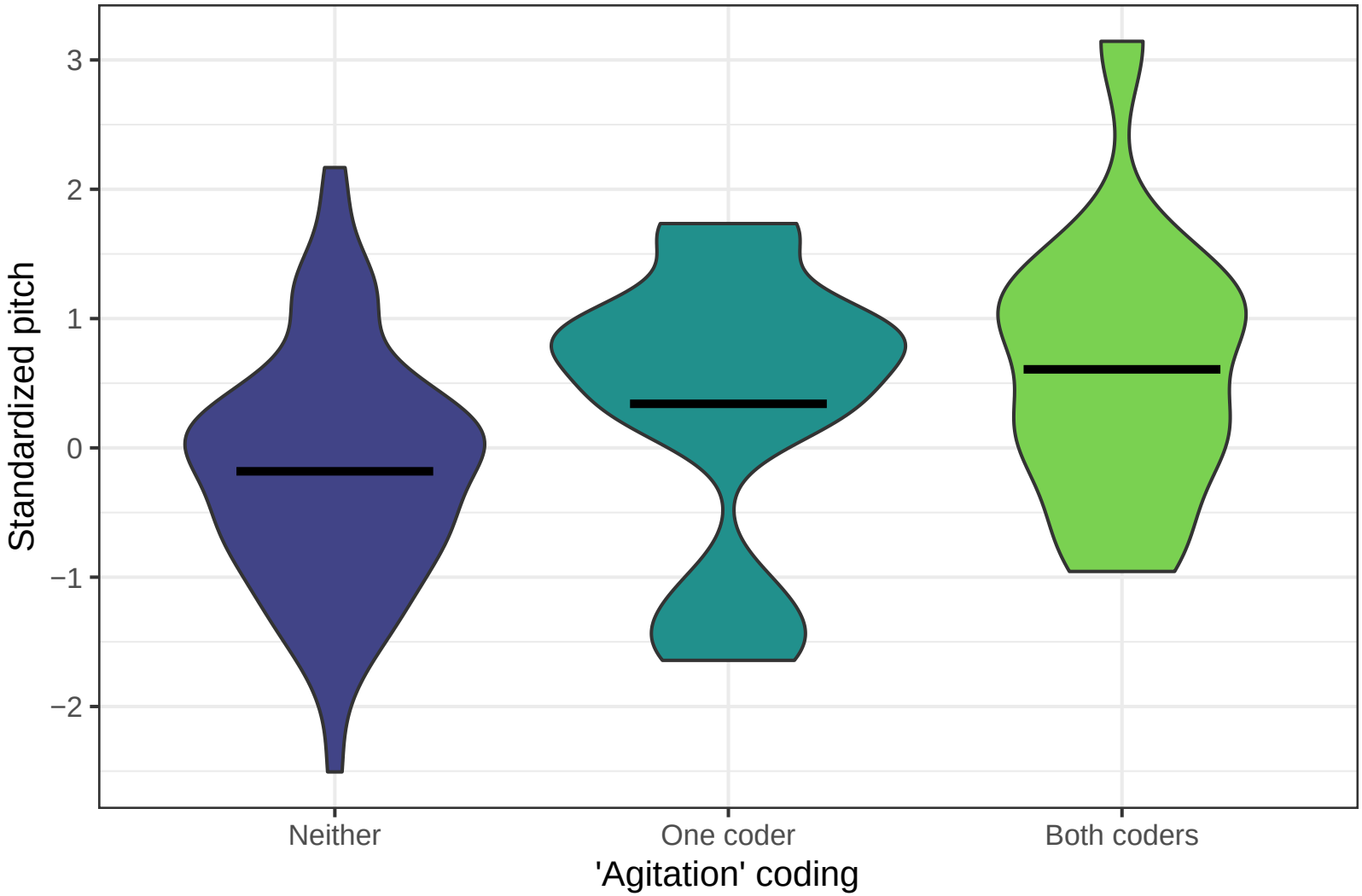
Manual annotation of 100 speeches for presence of agitation:



Validation I: Manual annotation

Back to presentation

Manual annotation of 100 speeches for presence of agitation:



$\rightsquigarrow$ coder agreement = 87 pct., Krippendorff's $\alpha = .62$

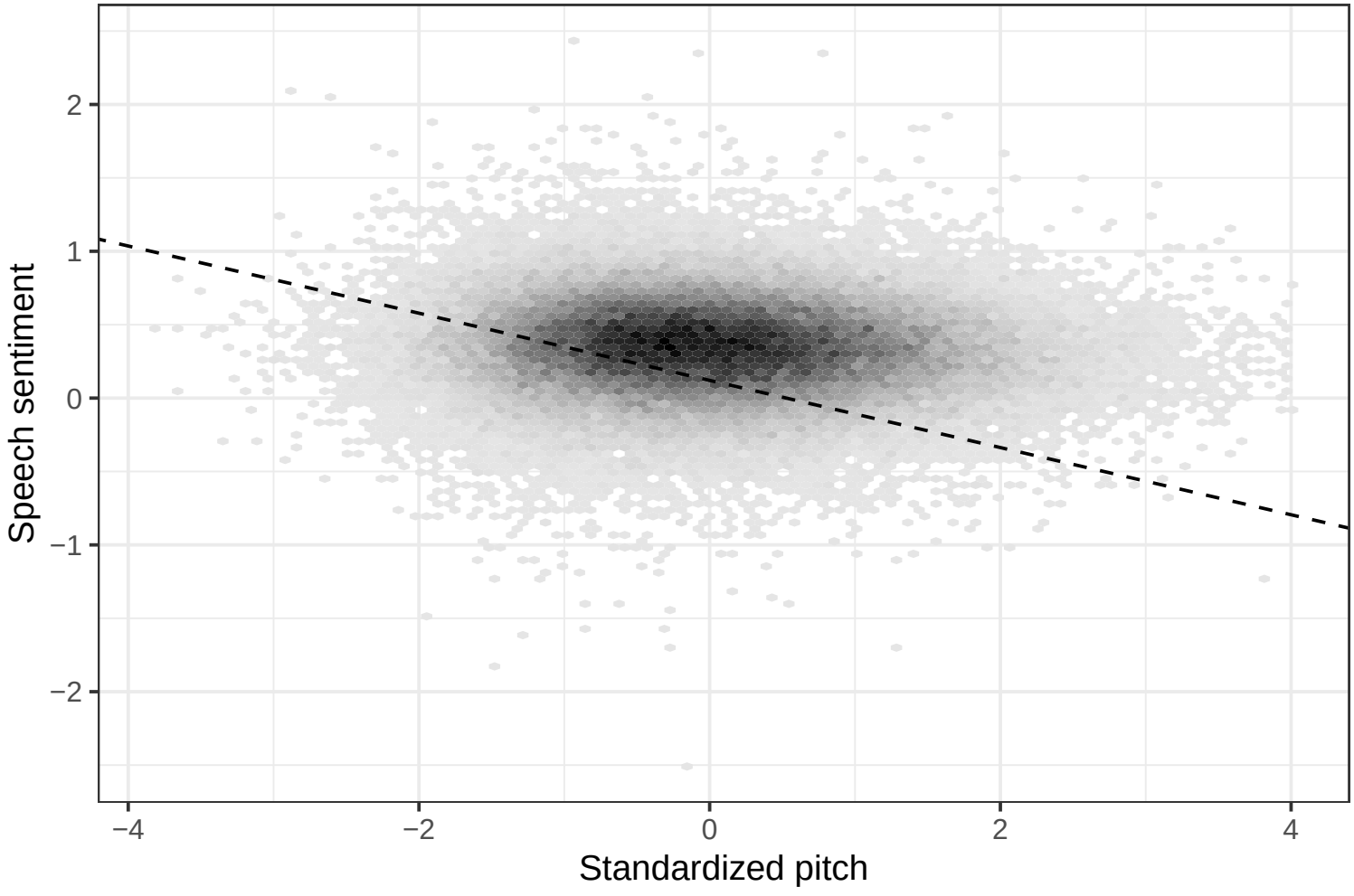
Validation II: Correlation w. text sentiment

Validation II: Correlation w. text sentiment

Standardized pitch vs. speech-level sentiment measure:

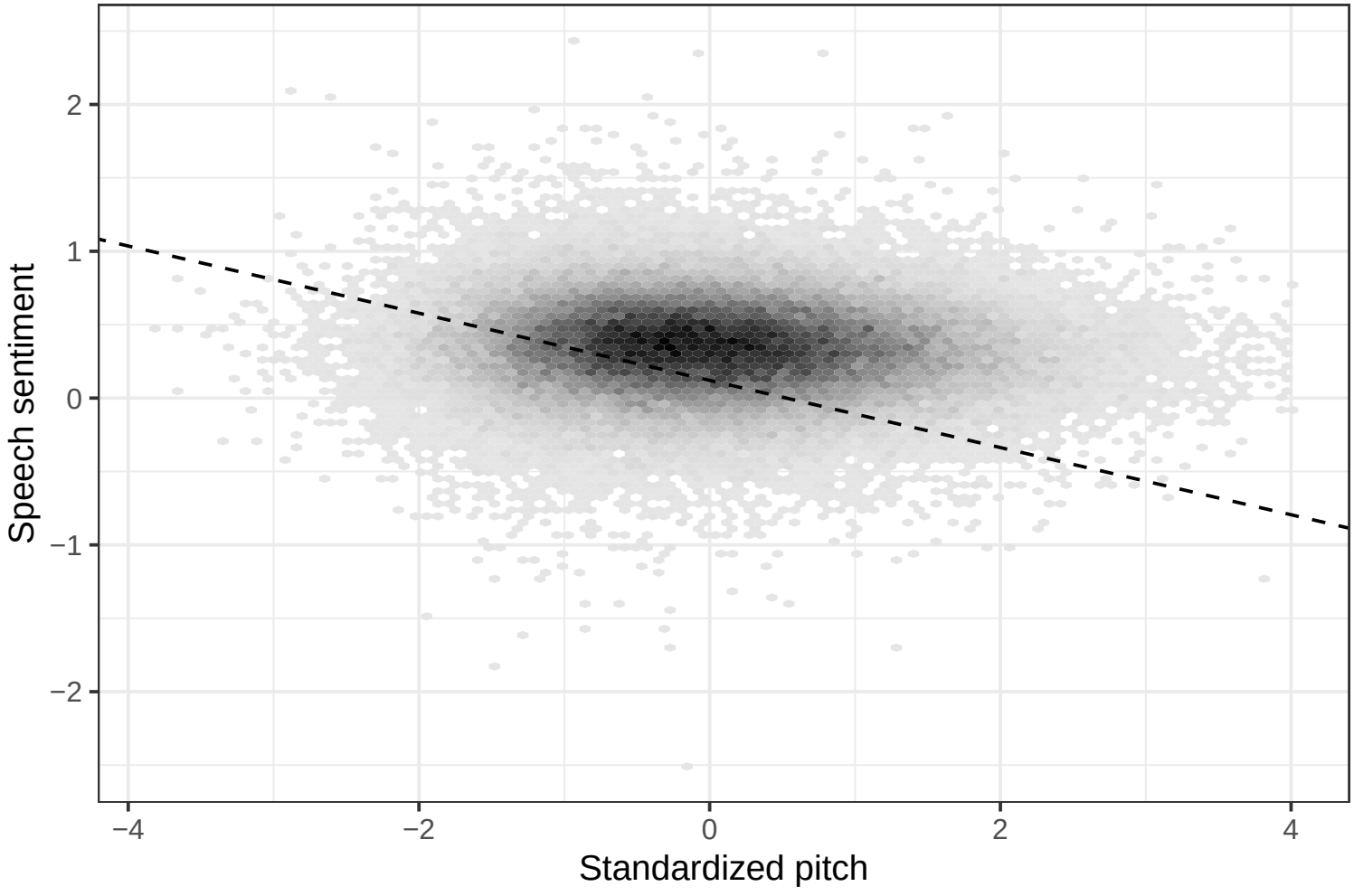
Validation II: Correlation w. text sentiment

Standardized pitch vs. speech-level sentiment measure:



Validation II: Correlation w. text sentiment

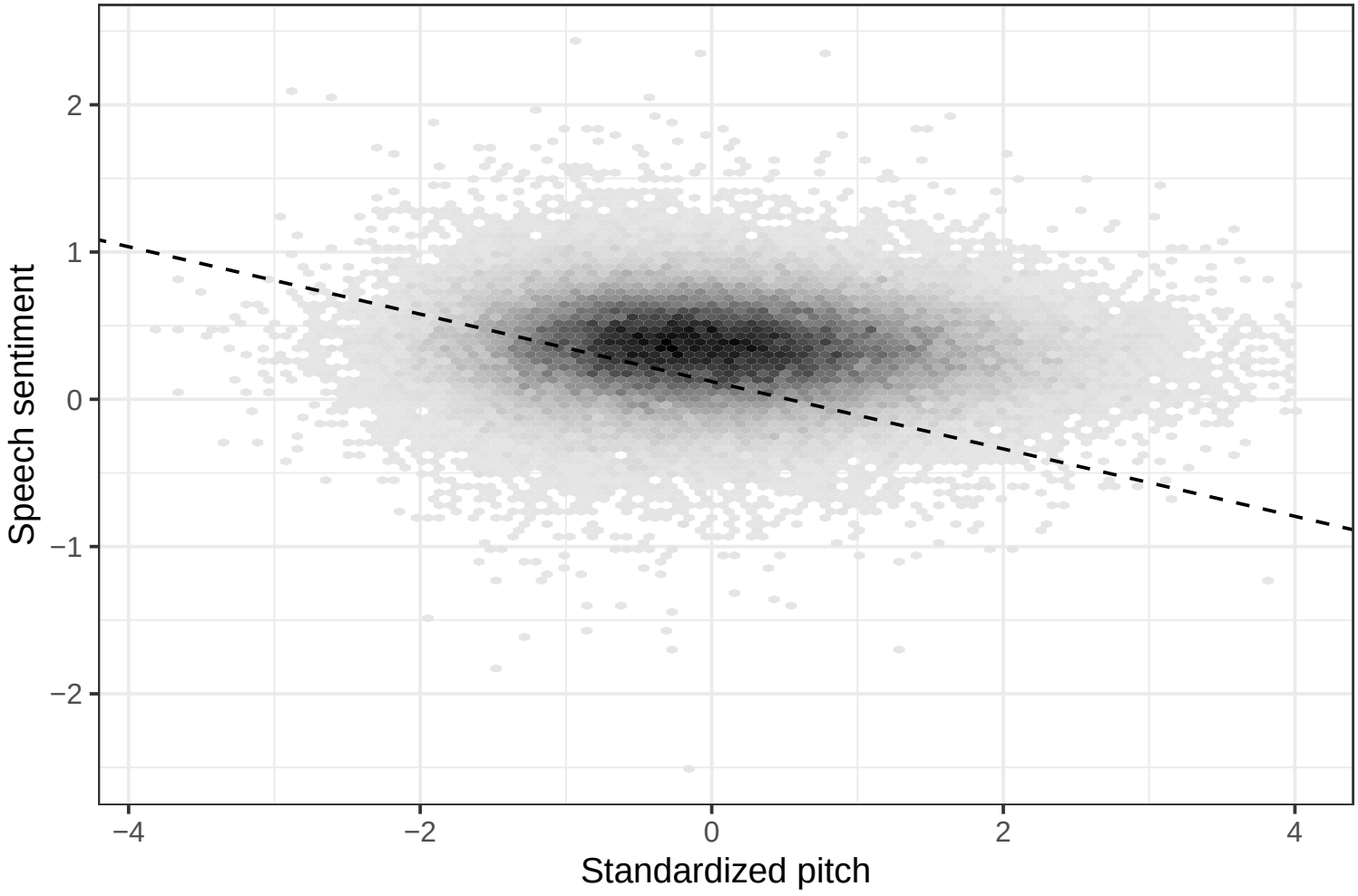
Standardized pitch vs. speech-level sentiment measure:



$\rightarrow$ speeches with higher std. pitch are more negative ($t = 19, p < .001$)

Validation II: Correlation w. text sentiment

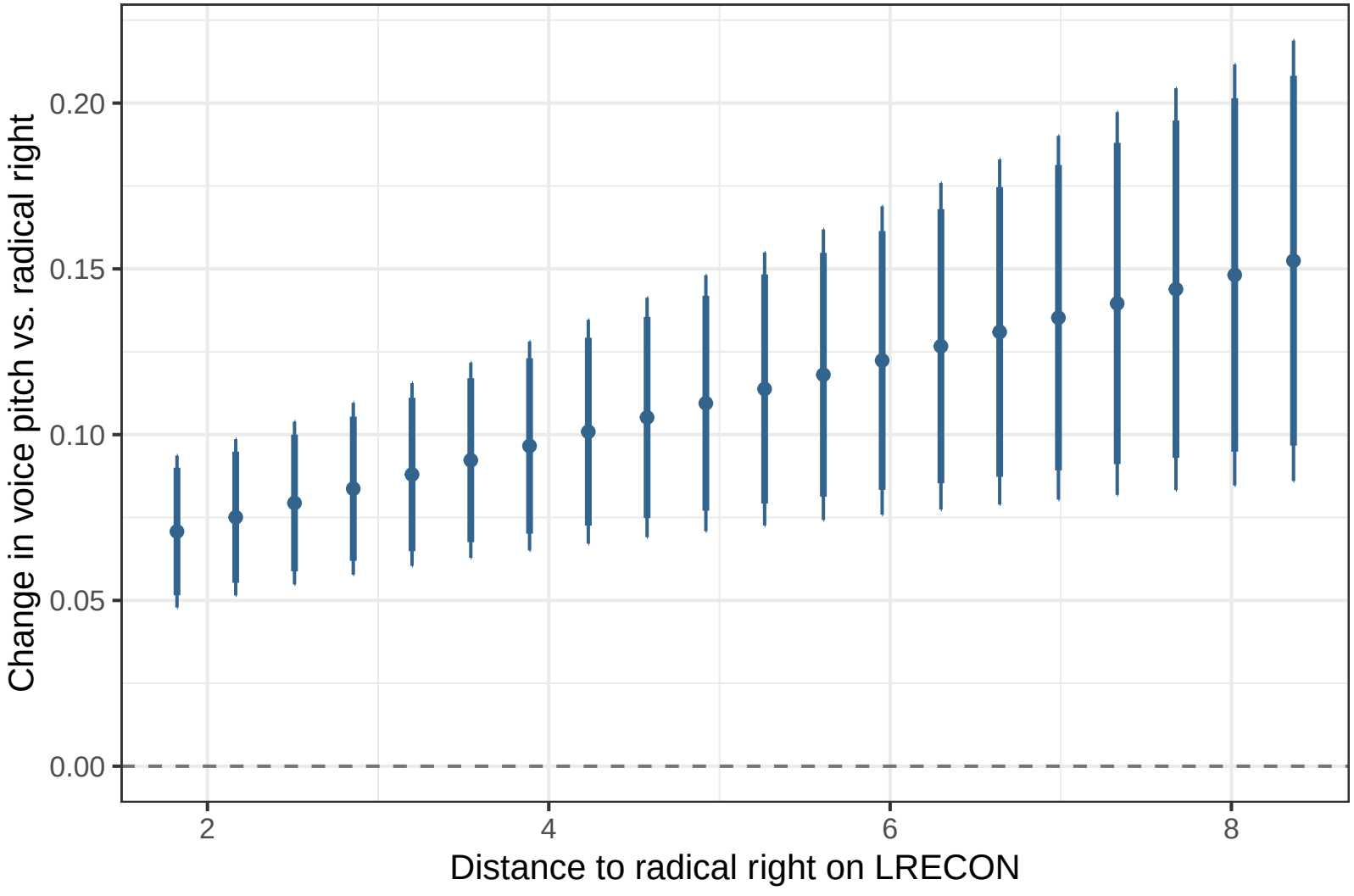
Standardized pitch vs. speech-level sentiment measure:



$\rightarrow$ speeches with higher std. pitch are more negative ($t = 19$, $p < .001$), correlation weaker in non-dyadic speeches

Other dimensions

Back to presentation



Other dimensions

Back to presentation