Elite Vocal Style and Mass Affective Polarization

Mathias Rask Aarhus University

Frederik Hjorth University of Copenhagen

Broader research interests

Broader research interests

Substantive

Conflict between notionally extreme and mainstream parties and voters

Broader research interests

Substantive

Conflict between notionally extreme and mainstream parties and voters

Methodological

Using tools and data from computational social science, e.g. unstructured data, machine learning

Broader research interests

Substantive

Conflict between notionally extreme and mainstream parties and voters

Methodological

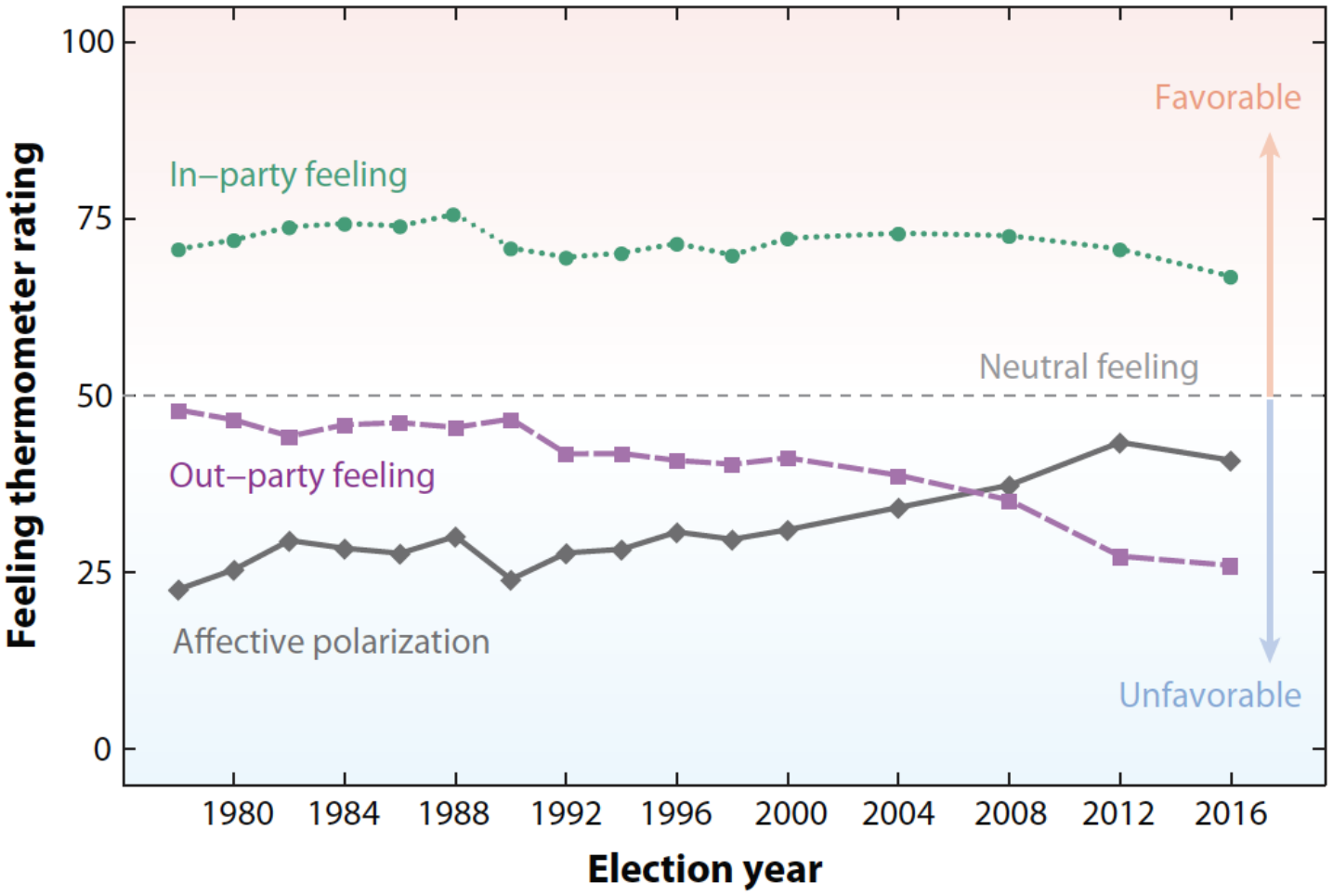
Using tools and data from computational social science, e.g. unstructured data, machine learning

Motivation

Why are some parties disliked more than others?

Source: https://www.theguardian.com/politics/2014/may/20/ukip-manifesto-europe-immigration

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.

What drives affective polarization?

What drives affective polarization?

- social sorting [9]

Source: https://www.scientificamerican.com/article/why-hatred-and-othering-of-political-foes-has-spiked-to-extreme-levels/

What drives affective polarization?

- social sorting [9]
- high-choice media environments [7]

Source: https://www.scientificamerican.com/article/why-hatred-and-othering-of-political-foes-has-spiked-to-extreme-levels/

What drives affective polarization?

- social sorting [9]
- high-choice media environments [7]
- personality traits [8]

Source: https://www.scientificamerican.com/article/why-hatred-and-othering-of-political-foes-has-spiked-to-extreme-levels/

What drives affective polarization?

- social sorting [9]
- high-choice media environments [7]
- personality traits [8]
- rising salience of ‘second-dimension’ issues [5]

Source: https://www.scientificamerican.com/article/why-hatred-and-othering-of-political-foes-has-spiked-to-extreme-levels/

What drives affective polarization?

- social sorting [9]
- high-choice media environments [7]
- personality traits [8]
- rising salience of 'second-dimension' issues [5]
- our focus here: **elite social interaction**

Source: https://www.scientificamerican.com/article/why-hatred-and-othering-of-political-foes-has-spiked-to-extreme-levels/

What drives affective polarization?

- social sorting [9]
- high-choice media environments [7]
- personality traits [8]
- rising salience of 'second-dimension' issues [5]
- our focus here: **elite social interaction**
- argument: elite social interaction conveys outparty dislike

Source: https://www.scientificamerican.com/article/why-hatred-and-othering-of-political-foes-has-spiked-to-extreme-levels/

Nonverbal aspects of elite interaction I

Nonverbal aspects of elite interaction I

- we focus on **nonverbal interaction**

Source: Andrew Harnik, AP

Nonverbal aspects of elite interaction I

Source: Andrew Harnik, AP

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

Nonverbal aspects of elite interaction I

Source: Andrew Harnik, AP

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

Nonverbal aspects of elite interaction I

Source: Andrew Harnik, AP

- we focus on **nonverbal interaction**
- essential part of interpersonal communication
- challenge existing literature's focus on verbal content
- $\rightsquigarrow$ existing evidence underestimates extent of elite partisan hostility

Nonverbal aspects of elite interaction II

Nonverbal aspects of elite interaction II

- focus here: vocal style

Nonverbal aspects of elite interaction II

- focus here: vocal style
- legislators use vocal style to express indignation at other parties, signaling target party extremity

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 aspects of elite interaction II

- focus here: vocal style
- legislators use vocal style to express indignation at other parties, signaling target party extremity
- 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

Hypothesis 1

Legislators express more indignation when addressing outbloc (vs. inbloc) legislators

Hypotheses

Hypothesis 1

Legislators express more indignation when addressing outbloc (vs. inbloc) legislators

Hypothesis 2

The effect of outbloc target is stronger for speeches on second-dimension issues

Hypotheses

Hypothesis 1

Legislators express more indignation when addressing outbloc (vs. inbloc) legislators

Hypothesis 2

The effect of outbloc target is stronger for speeches on second-dimension issues

Hypothesis 3

Between-party indignation is positively associated with between-party affective polarization among voters.

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 $\approx$ 5 hours
- remove recordings with only chairperson speaking
- extract audio channel from video using `ffmpeg`

Measurement I: Matching audio to speakers

Measurement I: Matching audio to speakers

Use `polannotate`, developed in separate working paper [10]:

Measurement I: Matching audio to speakers

Use `polannotate`, developed in separate working paper [10]:

1. use unsupervised learning to segment audio files into speech segments

Measurement I: Matching audio to speakers

Use `polannotate`, developed in separate working paper [10]:

1. use unsupervised learning to segment audio files into speech segments
2. use automatic speech recognition on each audio segment to generate hypothesis text

Measurement I: Matching audio to speakers

Use **`polannotate`**, developed in separate working paper [10]:

1. use unsupervised learning to segment audio files into speech segments
2. use automatic speech recognition on each audio segment to generate hypothesis text
3. generate embeddings for each audio segment to encode speaker identity

Measurement I: Matching audio to speakers

Use `polannotate`, developed in separate working paper [10]:

1. use unsupervised learning to segment audio files into speech segments
2. use automatic speech recognition on each audio segment to generate hypothesis text
3. generate embeddings for each audio segment to encode speaker identity
4. using ②+③ as supervisory signals, match audio segment to transcript speeches and speakers

Measurement I: Matching audio to speakers

Use `polannotate`, developed in separate working paper [10]:

1. use unsupervised learning to segment audio files into speech segments
2. use automatic speech recognition on each audio segment to generate hypothesis text
3. generate embeddings for each audio segment to encode speaker identity
4. using ②+③ as supervisory signals, match audio segment to transcript speeches and speakers

Measurement I: Matching audio to speakers

Use **`polannotate`**, developed in separate working paper [10]:

1. use unsupervised learning to segment audio files into speech segments
2. use automatic speech recognition on each audio segment to generate hypothesis text
3. generate embeddings for each audio segment to encode speaker identity
4. using ②+③ as supervisory signals, match audio segment to transcript speeches and speakers

→ linked audio and text for ≈ 209,000 speeches

Measurement I: Matching audio to speakers

Use **`polannotate`**, developed in separate working paper [10]:

1. use unsupervised learning to segment audio files into speech segments
2. use automatic speech recognition on each audio segment to generate hypothesis text
3. generate embeddings for each audio segment to encode speaker identity
4. using ②+③ as supervisory signals, match audio segment to transcript speeches and speakers

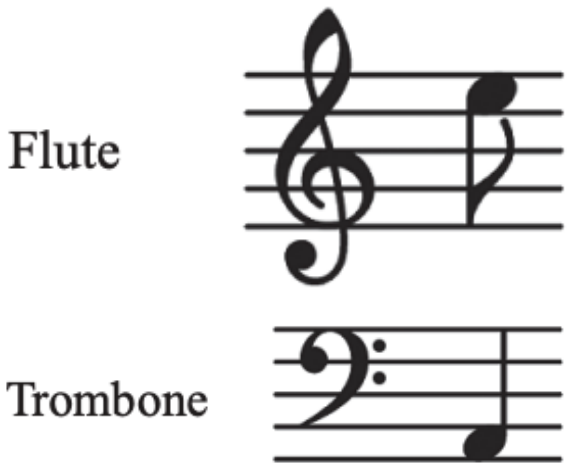
→ linked audio and text for ≈ 209,000 speeches

▸ Additional detail

Measurement II: Measuring indignation

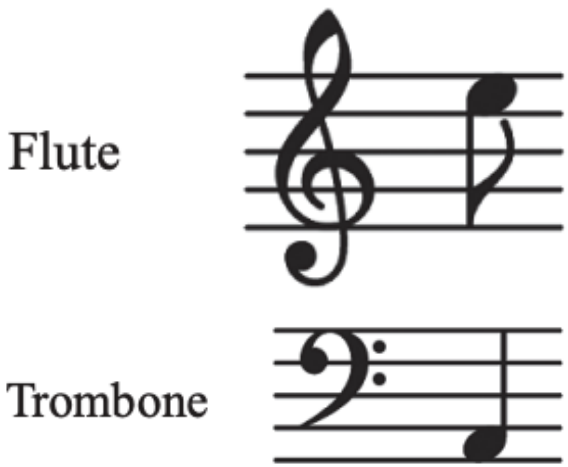
Measurement II: Measuring indignation

- use standardized vocal pitch as proxy for indignation



High vs. low pitch illustration from Knox & Lucas [6]

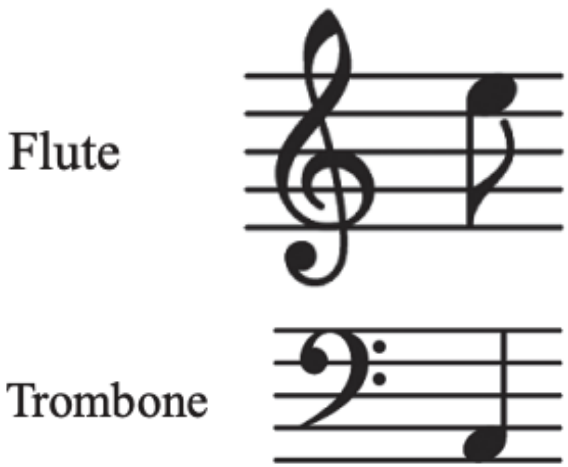
Measurement II: Measuring indignation



High vs. low pitch illustration from Knox & Lucas [6]

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

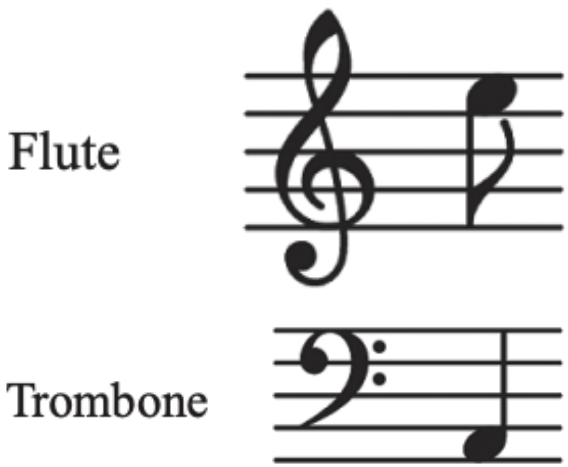
Measurement II: Measuring indignation



High vs. low pitch illustration from Knox & Lucas [6]

- use standardized vocal pitch as proxy for indignation
- 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)

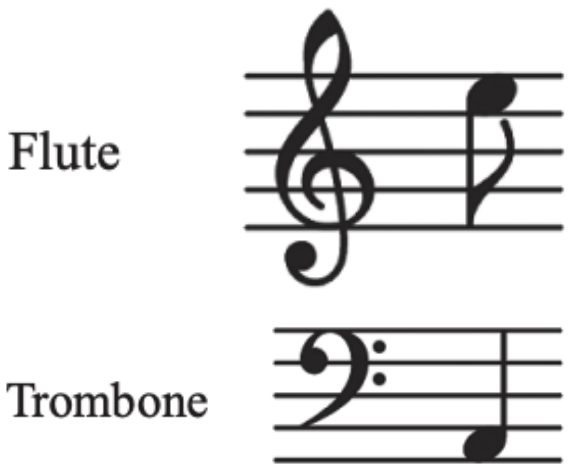
Measurement II: Measuring indignation



High vs. low pitch illustration from Knox & Lucas [6]

- use standardized vocal pitch as proxy for indignation
- 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

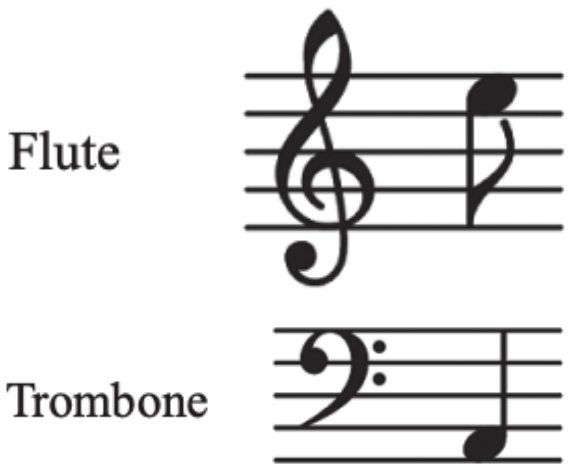
Measurement II: Measuring indignation



High vs. low pitch illustration from Knox & Lucas [6]

- use standardized vocal pitch as proxy for indignation
- 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 II: Measuring indignation



High vs. low pitch illustration from Knox & Lucas [6]

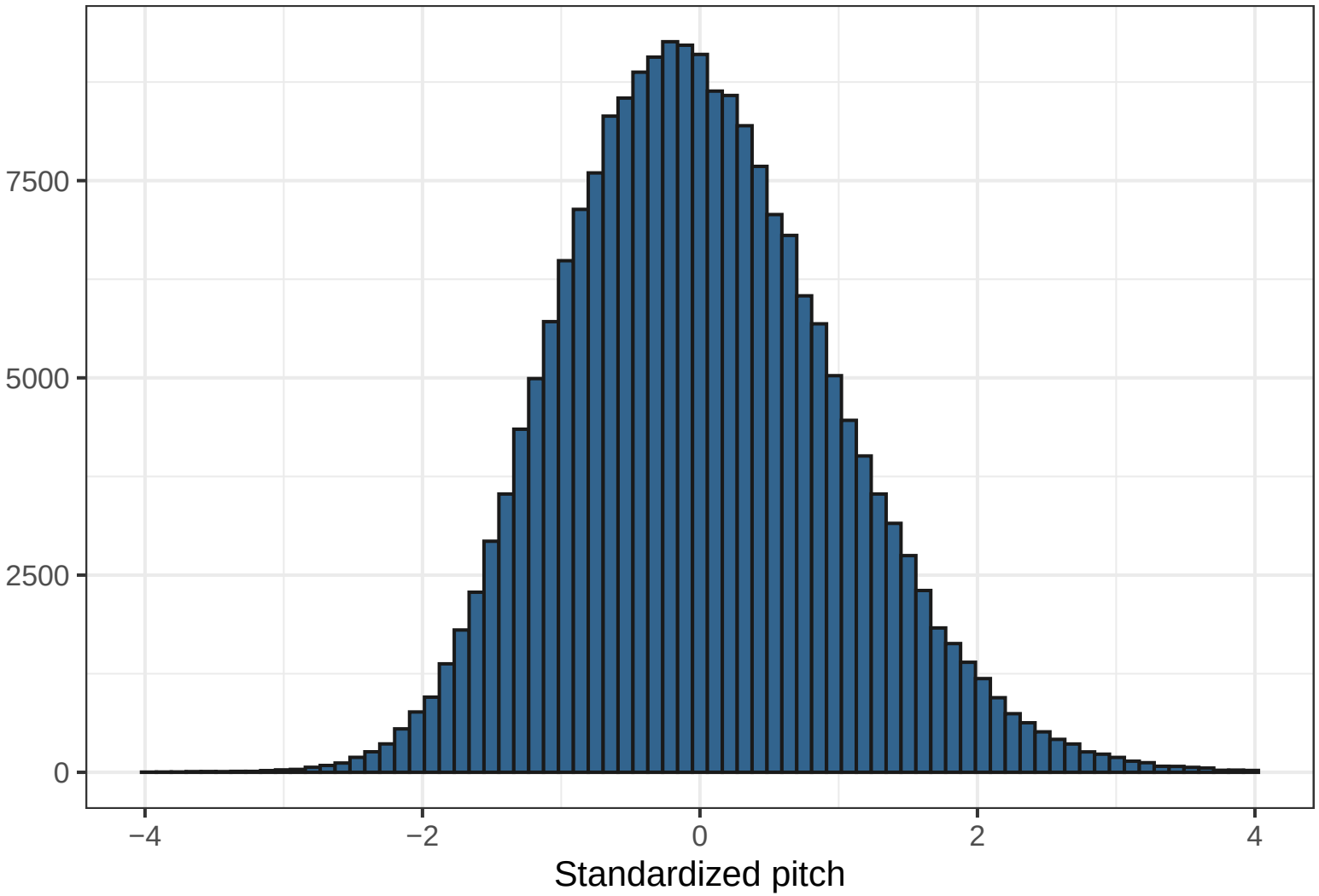
- use standardized vocal pitch as proxy for indignation
- context-dependent meaning: pitch used in earlier work to measure issue commitment [4] and judicial intent [3]
- standardize pitch within each speaker to remove speaker heterogeneity (e.g. gender)
- $\rightsquigarrow$ implicit control for all fixed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

▸ Additional detail

▸ Low pitch example

▸ High pitch example

Distribution of standardized pitch



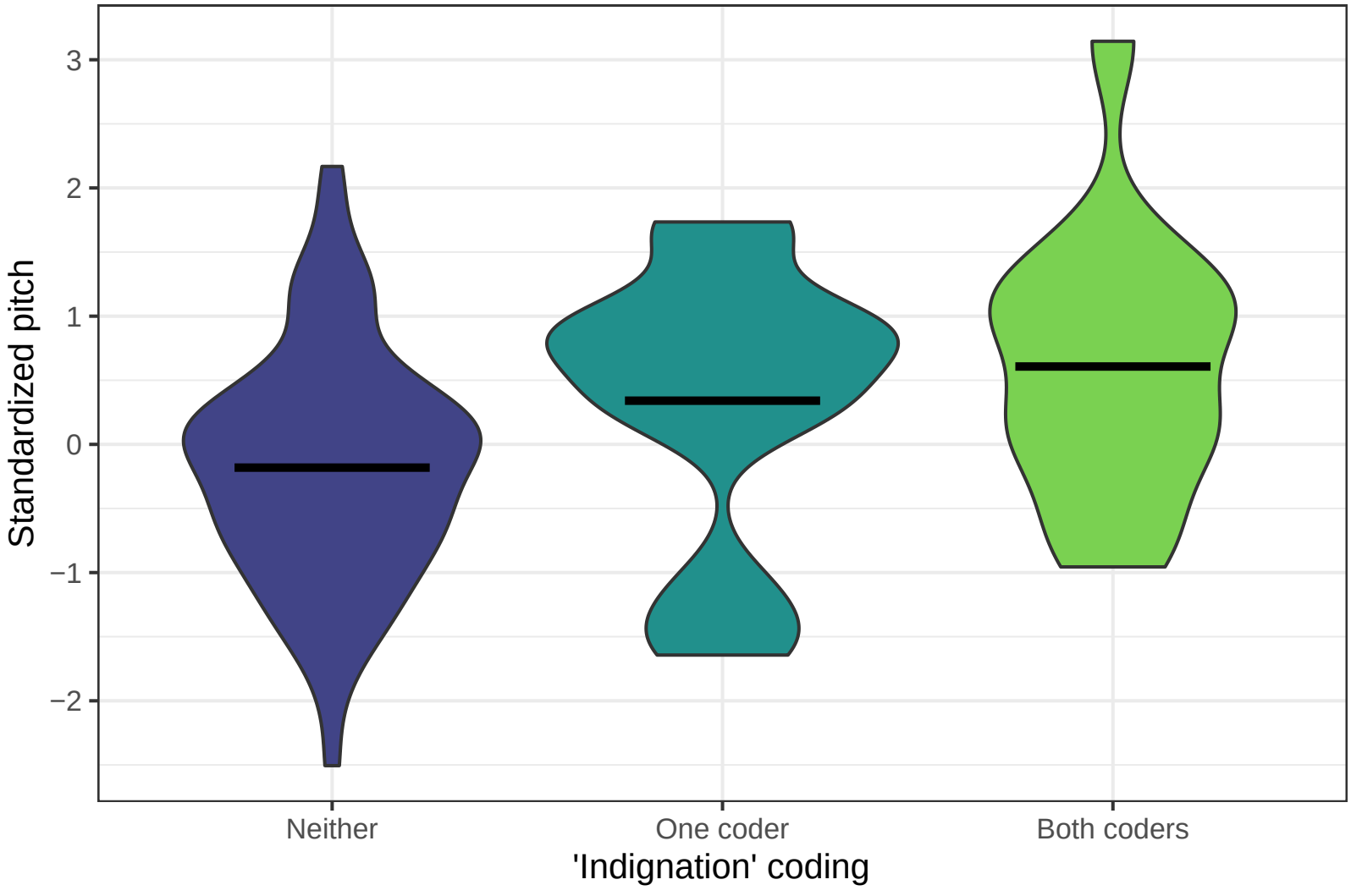
Validation I: Manual annotation

Validation I: Manual annotation

Manual annotation of 100 speeches for presence of indignation:

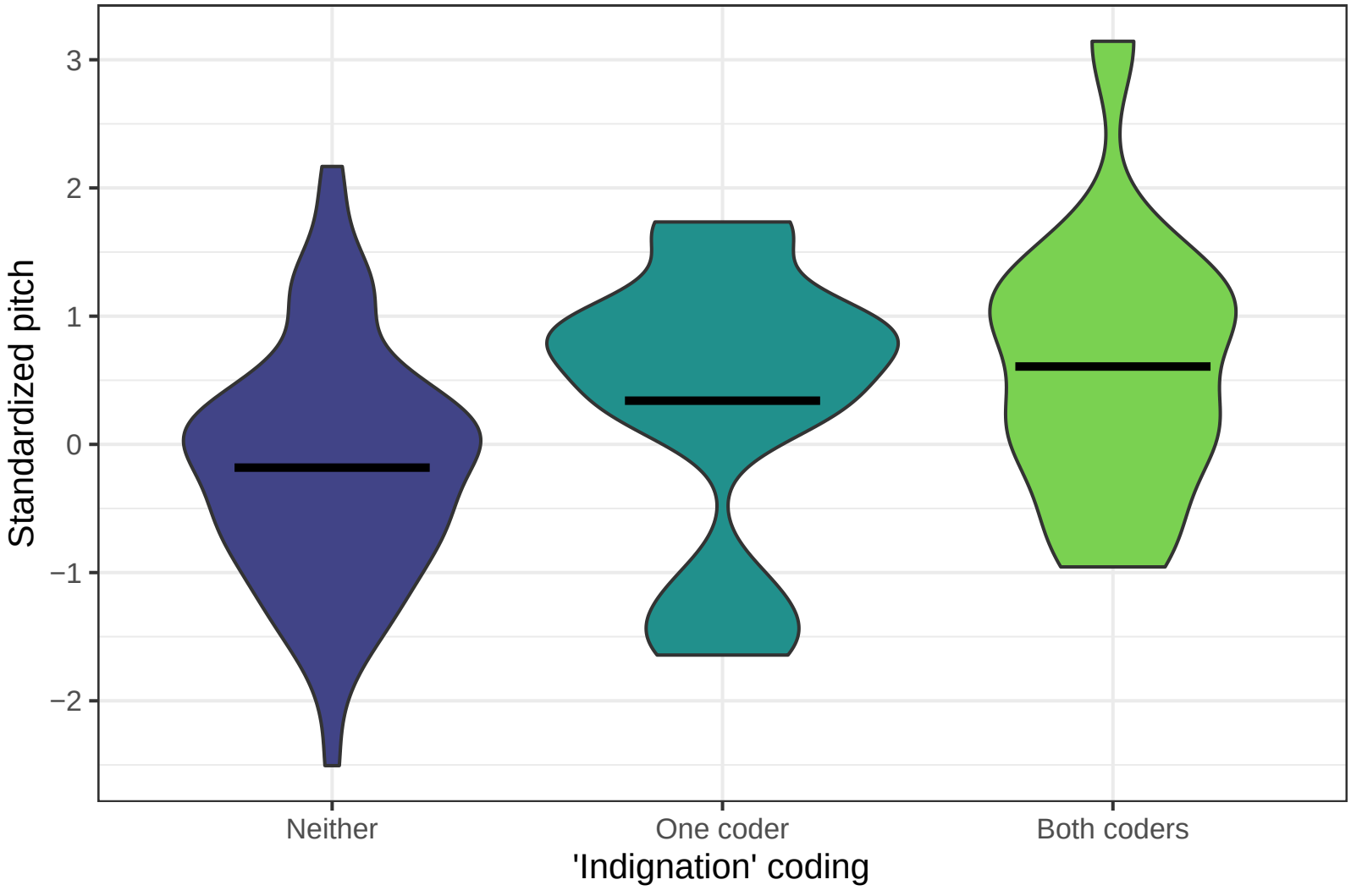
Validation I: Manual annotation

Manual annotation of 100 speeches for presence of indignation:



Validation I: Manual annotation

Manual annotation of 100 speeches for presence of indignation:



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

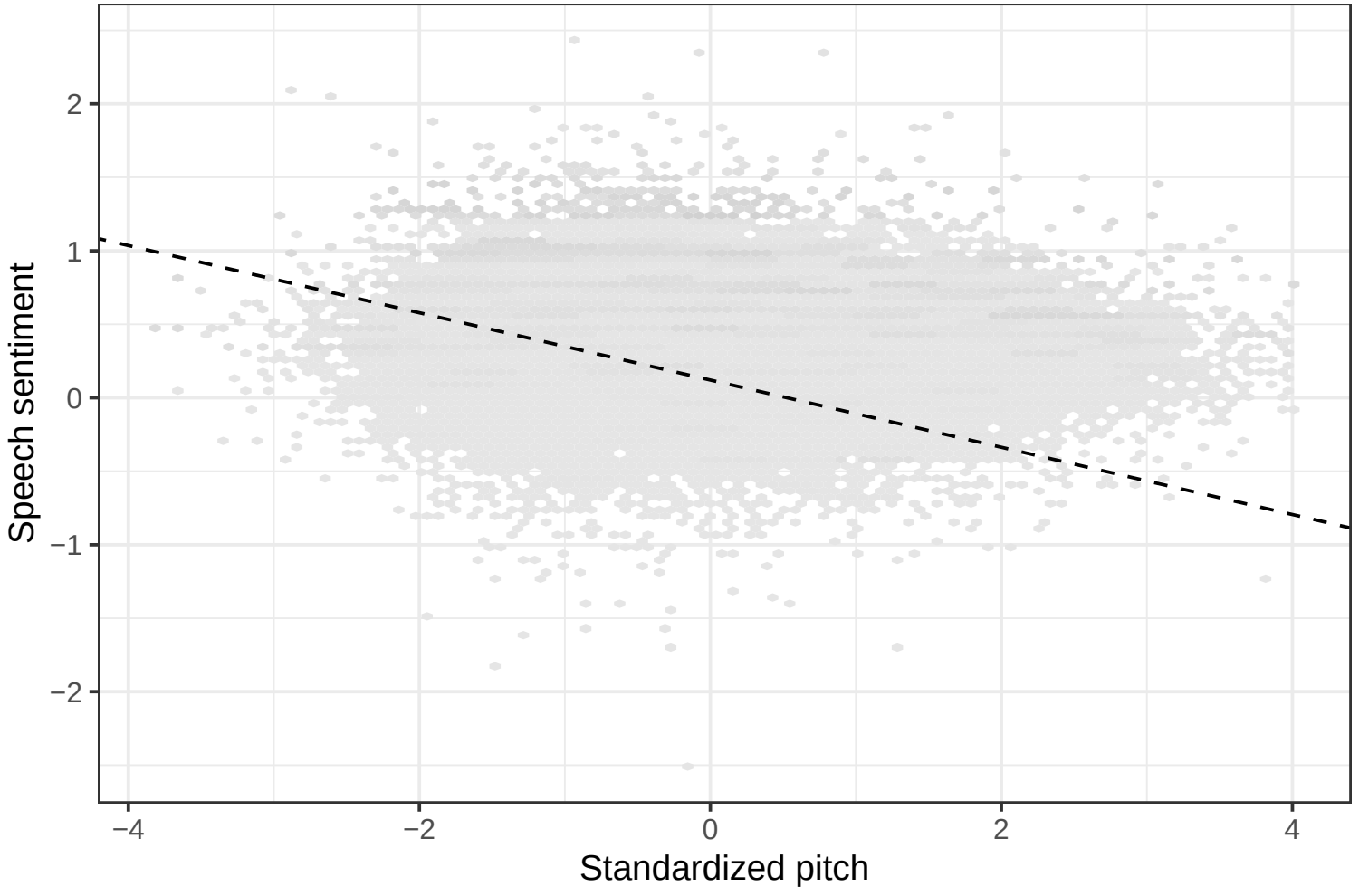
Validation II: Correlation w. text sentiment

Validation II: Correlation w. text sentiment

Standardized pitch vs. speech-level sentiment measure:

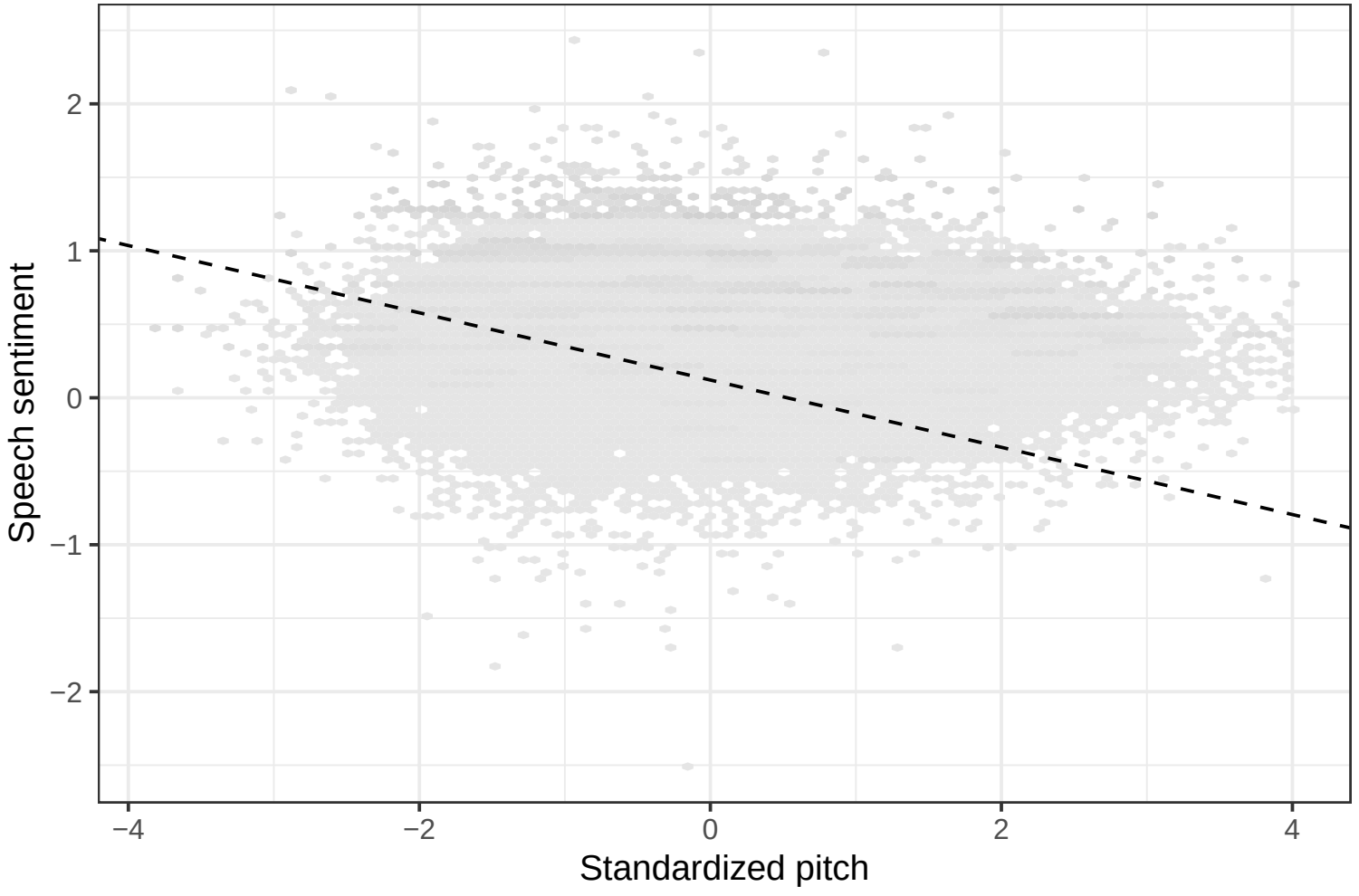
Validation II: Correlation w. text sentiment

Standardized pitch vs. speech-level sentiment measure:



Validation II: Correlation w. text sentiment

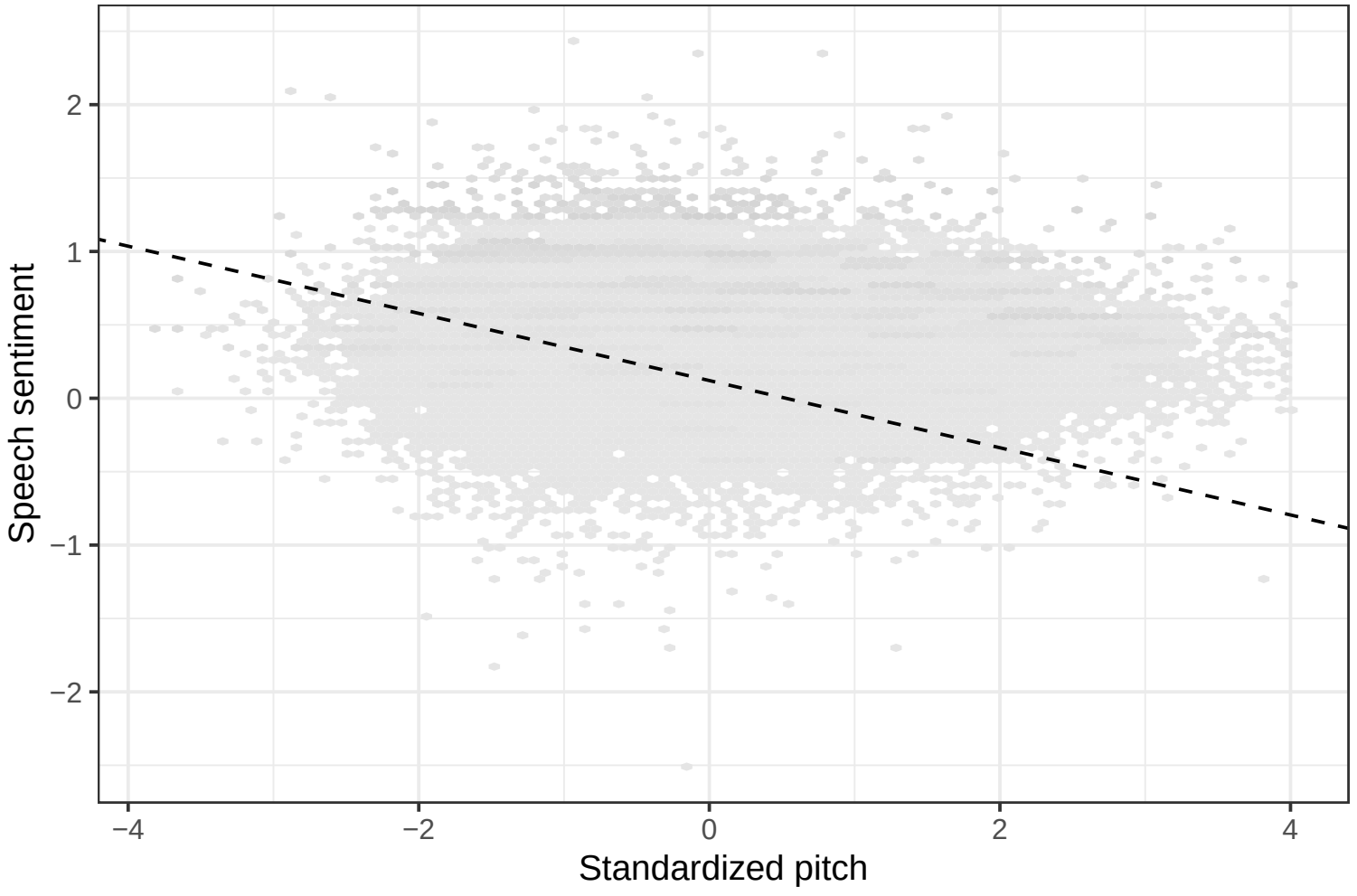
Standardized pitch vs. speech-level sentiment measure:



$\rightarrow$ 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

Measurement III: Measuring second dimension alignment

Measurement III: Measuring second dimension alignment

- goal: scale speeches ranging from maximally first dimension (e.g., economic redistribution) to maximally second dimension (e.g., immigration)

Measurement III: Measuring second dimension alignment

- goal: scale speeches ranging from maximally first dimension (e.g., economic redistribution) to maximally second dimension (e.g., immigration)
- method: latent semantic scaling [12]

Measurement III: Measuring second dimension alignment

- goal: scale speeches ranging from maximally first dimension (e.g., economic redistribution) to maximally second dimension (e.g., immigration)
- method: latent semantic scaling [12]
- $\rightsquigarrow$ uses word embeddings to estimate 'polarity' of words based on user-defined seed word lists for each scale endpoint

Measurement III: Measuring second dimension alignment

- goal: scale speeches ranging from maximally first dimension (e.g., economic redistribution) to maximally second dimension (e.g., immigration)
- method: latent semantic scaling [12]
- $\rightsquigarrow$ uses word embeddings to estimate 'polarity' of words based on user-defined seed word lists for each scale endpoint
- we scale speeches based on first and second dimension seed word lists (e.g. `taxes`, `financial` vs. `immigration`, `moral`)

Measurement IV: Identifying party dyads

Measurement IV: 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 IV: 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 IV: 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

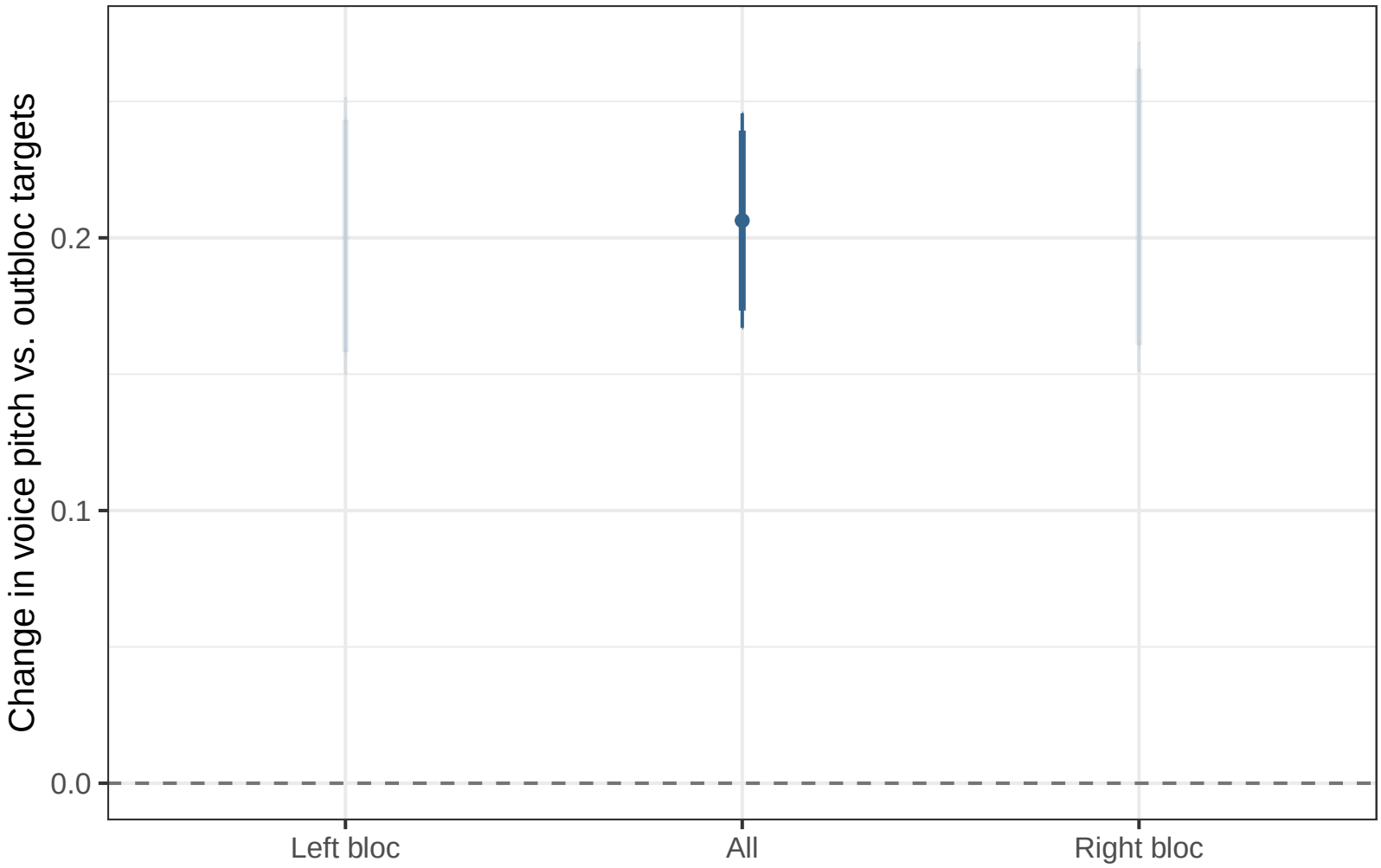
Measurement IV: Identifying party dyads

- 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 H3, compare to voter dyadic affect measured in election surveys

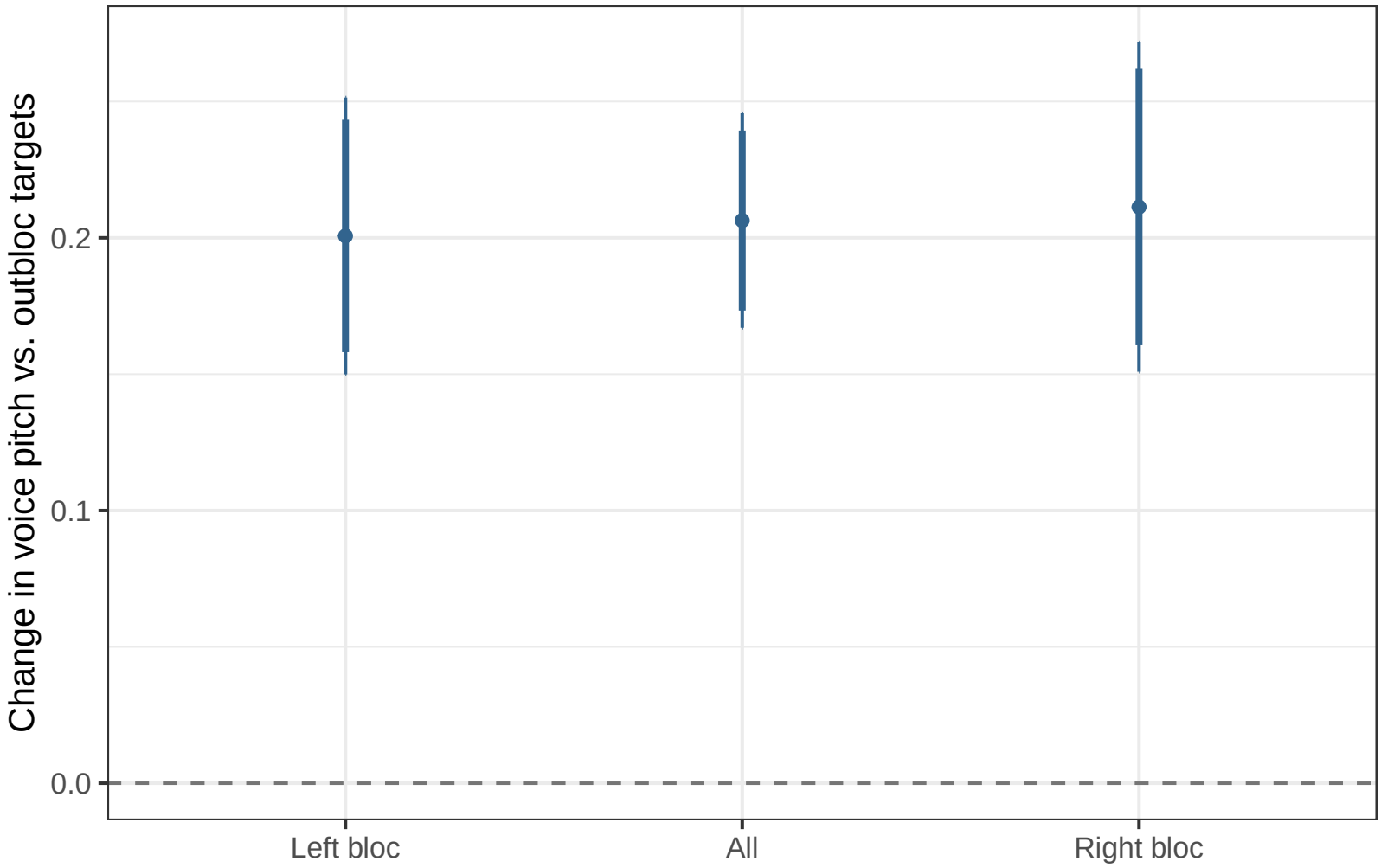
Results

1: Indignation vs. outbloc targets

1: Indignation vs. outbloc targets

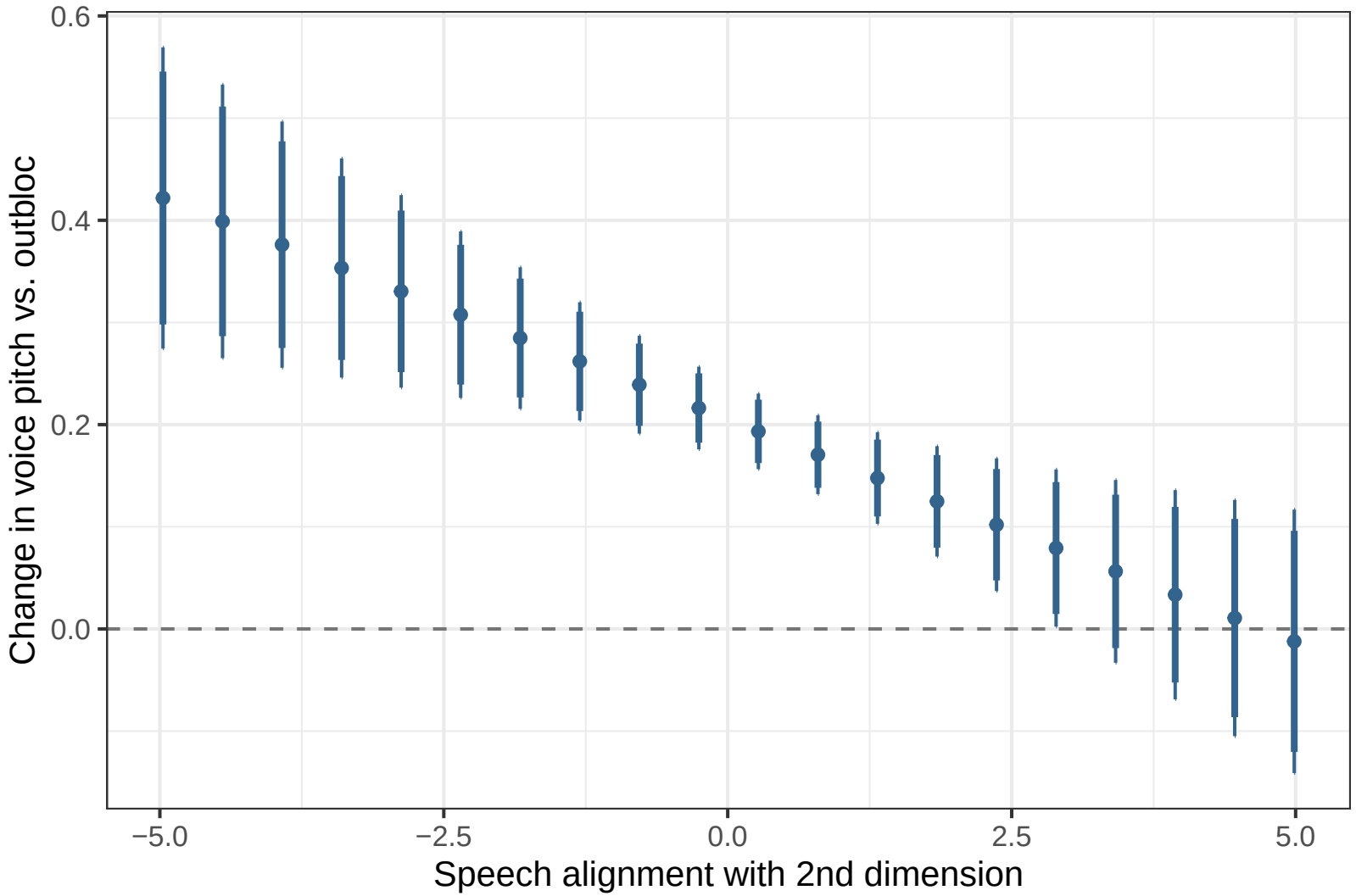


1: Indignation vs. outbloc targets



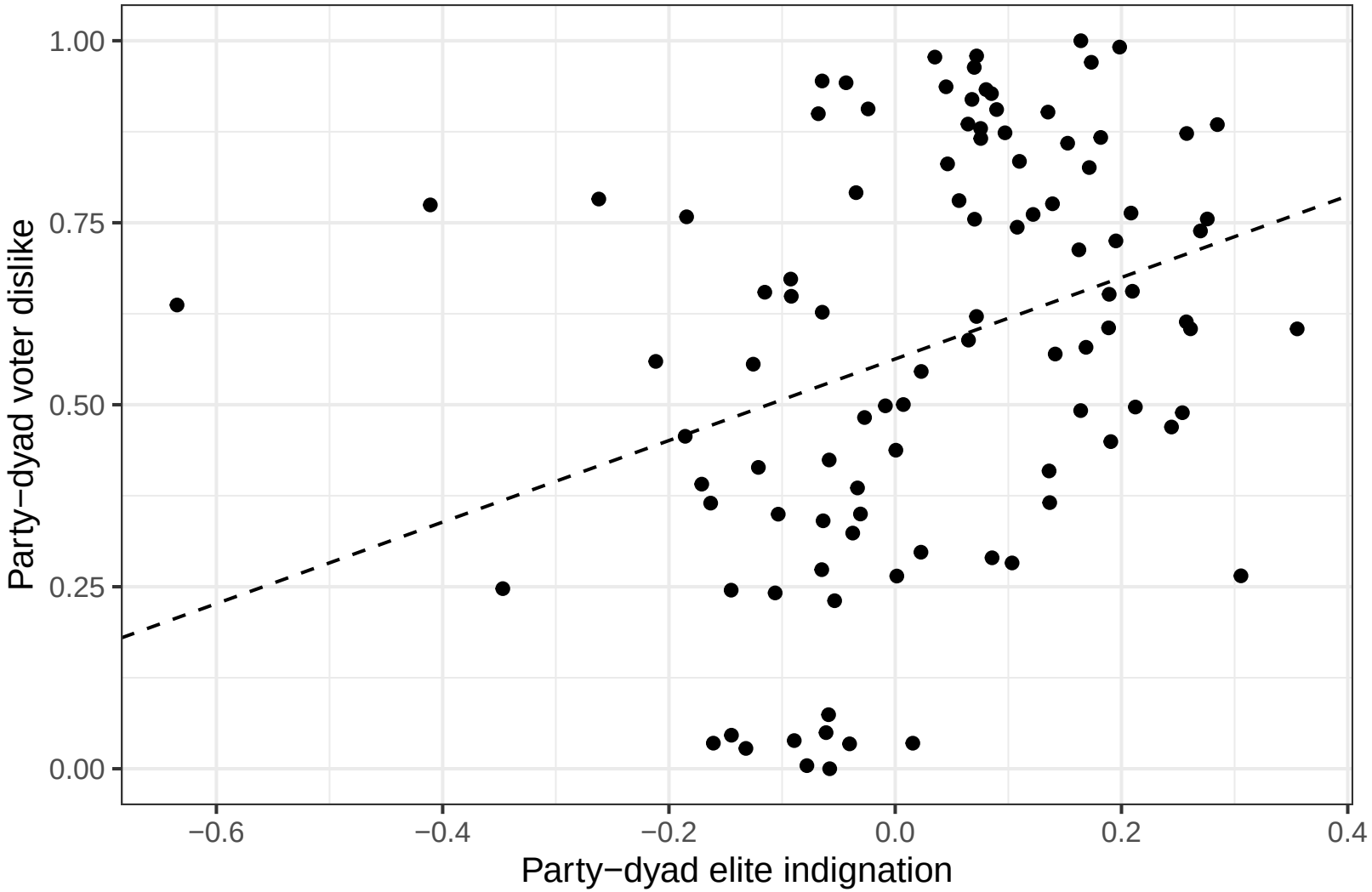
2: Indignation change vs. second dimension alignment

2: Indignation change vs. second dimension alignment

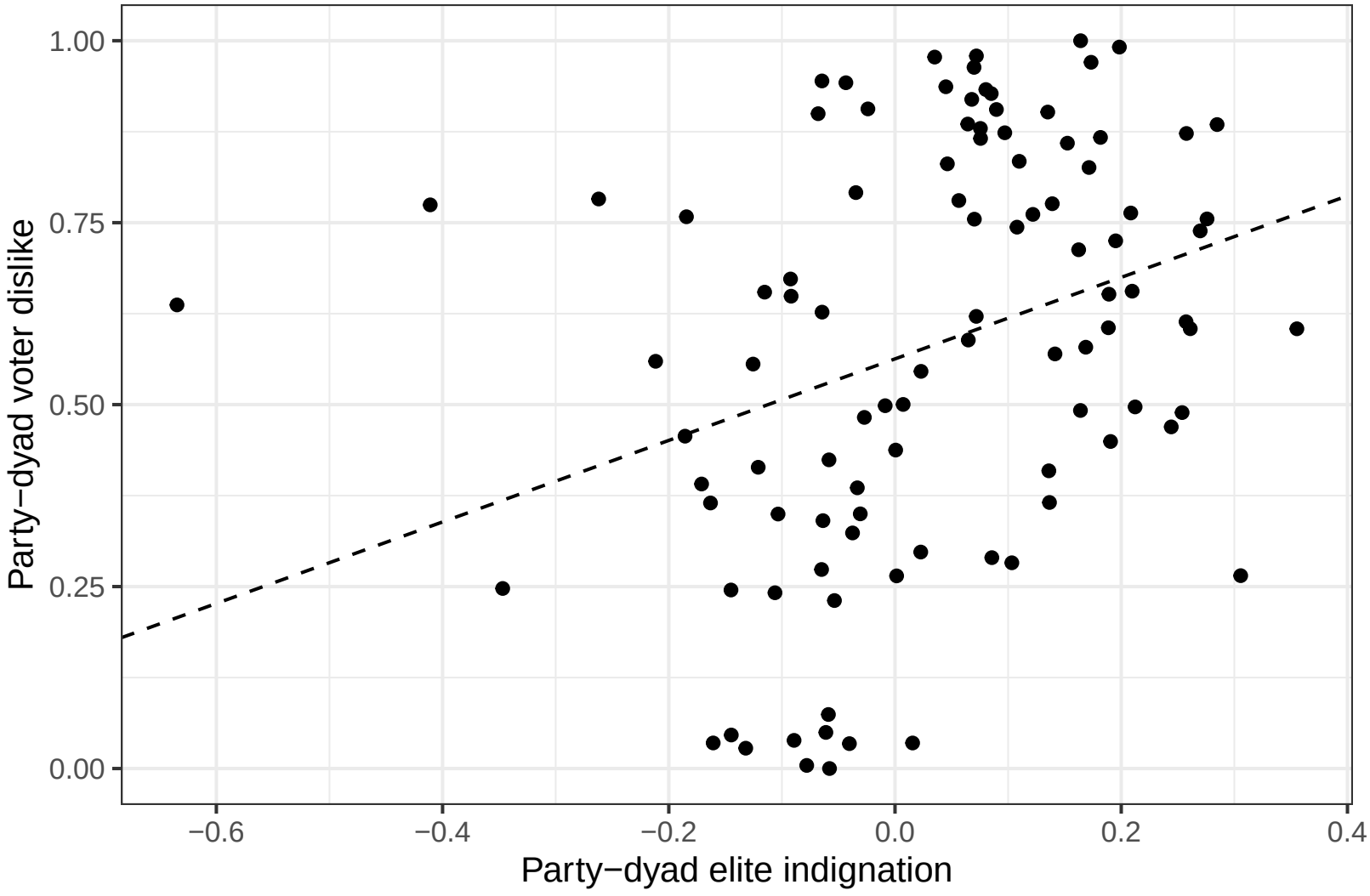


3: Party-dyad indignation vs. party-dyad voter dislike

3: Party-dyad indignation vs. party-dyad voter dislike



3: Party-dyad indignation vs. party-dyad voter dislike



☞ robust to speaker party FEs, exclusion of radical right targets

Conclusion and Implications

Conclusion and Implications

Conclusion and Implications

- we study vocal style in parliamentary speech, focusing on indignation

Conclusion and Implications

- we study vocal style in parliamentary speech, focusing on indignation
  1. higher when target is from opposite bloc

Conclusion and Implications

- we study vocal style in parliamentary speech, focusing on indignation
  1. higher when target is from opposite bloc
  2. higher for outbloc targets mainly on first-dimension issues

Conclusion and Implications

- we study vocal style in parliamentary speech, focusing on indignation
  1. higher when target is from opposite bloc
  2. higher for outbloc targets mainly on first-dimension issues
  3. correlated with voter dyadic affect

Conclusion and Implications

- we study vocal style in parliamentary speech, focusing on indignation
  1. higher when target is from opposite bloc
  2. higher for outbloc targets mainly on first-dimension issues
  3. correlated with voter dyadic affect
- demonstrates elites use vocal style to signal partisan dislike

Conclusion and Implications

- we study vocal style in parliamentary speech, focusing on indignation
  1. higher when target is from opposite bloc
  2. higher for outbloc targets mainly on first-dimension issues
  3. correlated with voter dyadic affect
- demonstrates elites use vocal style to signal partisan dislike
- implies content-only approaches underestimate expressed partisan hostility

Conclusion and Implications

- we study vocal style in parliamentary speech, focusing on indignation
  1. higher when target is from opposite bloc
  2. higher for outbloc targets mainly on first-dimension issues
  3. correlated with voter dyadic affect
- demonstrates elites use vocal style to signal partisan dislike
- implies content-only approaches underestimate expressed partisan hostility
- consistent with role of elite social interaction in sustaining mass affective polarization

Thanks for your attention!

Elite Vocal Style and Mass Affective Polarization

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.


N. Gidron, J. Adams, and W. Horne.
Who dislikes whom? affective polarization between pairs of parties in western democracies.
British Journal of Political Science, page 1–19, 2022.


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

References iii

Y. Lelkes, G. Sood, and S. Iyengar.
The hostile audience: The effect of access to broadband internet on partisan affect.
American Journal of Political Science, 61(1):5–20, 2017.

M. D. Luttig.
The "prejudiced personality" and the origins of partisan strength, affective polarization, and partisan sorting.
Political Psychology, 39:239–256, 2018.

L. Mason.
***Uncivil agreement: How politics became our identity*.**
University of Chicago Press, 2018.

M. Rask.
Polannotate: Matching audio to transcripts.
Working Paper, 2023.

References iv


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.


K. Watanabe.
Latent semantic scaling: A semisupervised text analysis technique for new domains and languages.
Communication Methods and Measures, 15(2):81–102, 2021.

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 [11]) 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 and speech recognition in step 2.

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

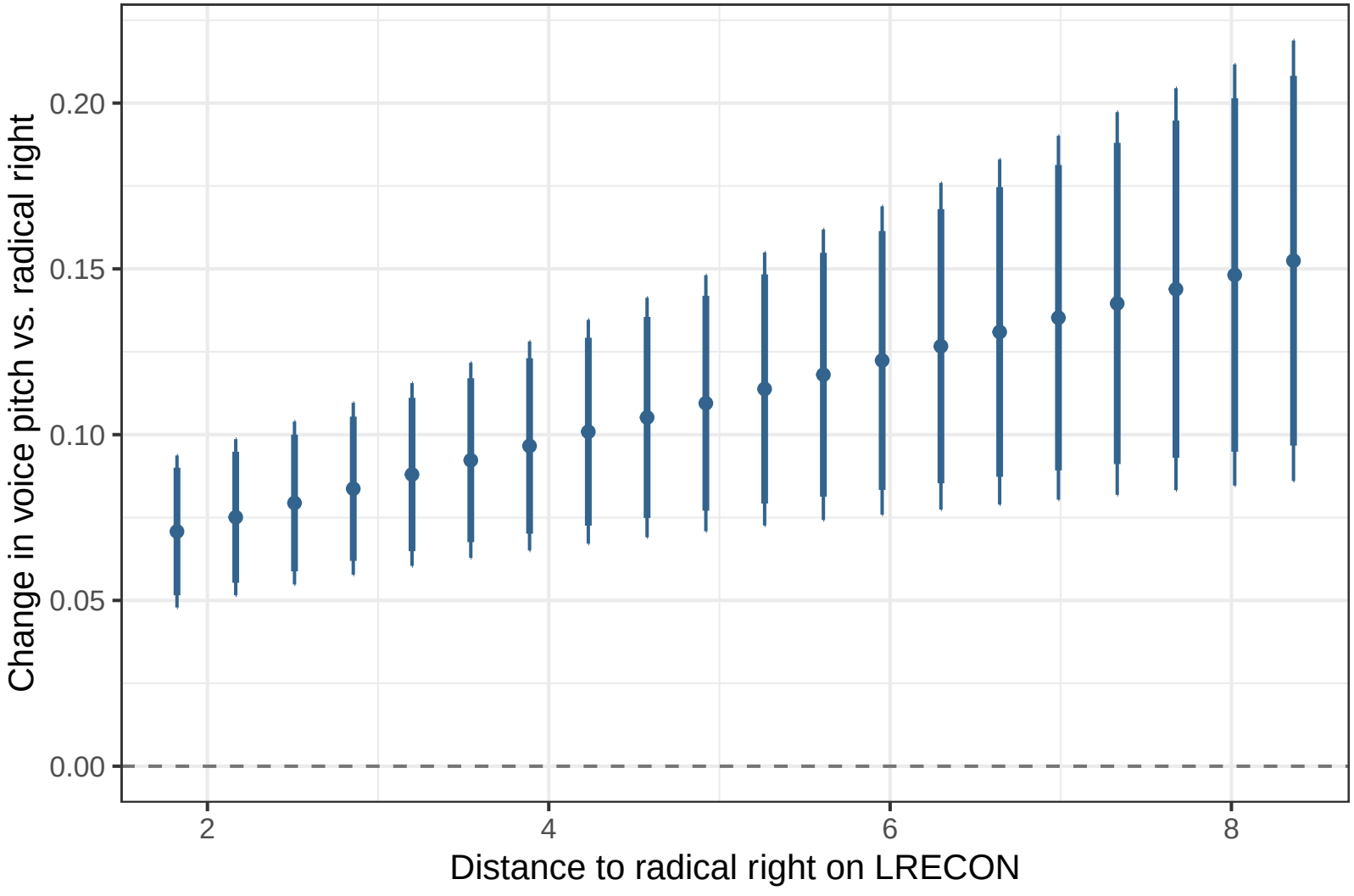
Back to presentation

Standardized vocal pitch as a proxy for indignation:

- estimate pitch for all speeches using open-source software `communication` (R package by [6])
  - 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)

Other dimensions

Back to presentation



Other dimensions

Back to presentation