British Journal of Political Science (2025), **55**, e165, 1–21
doi:10.1017/S0007123425100872

British Journal of Political Science

ARTICLE

Public Opinion and Emphatic Legislative Speech: Evidence from an Automated Video Analysis

Oliver Rittmann[1], Tobias Ringwald[2] and Dominic Nyhuis[3]

[1]Mannheim Centre for European Social Research (MZES), University of Mannheim, Mannheim, Germany, [2]Independent Researcher and [3]Institute for Political Science, Leibniz University Hannover, Hannover, Germany
Corresponding author: Oliver Rittmann; Email: oliver.rittmann@uni-mannheim.de

(Received 26 August 2024; revised 11 August 2025; accepted 12 August 2025)

Abstract
Why do politicians sometimes deliver passionate speeches and sometimes tedious monologues? Even though the delivery is key to understanding political speech, we know little about when and why political actors choose particular delivery styles. Focusing on legislative speech, we expect legislators to deliver more emphatic speeches when their vote is aligned with the preferences of their constituents. To test this proposition, we develop and apply an automated video analysis model to speech recordings from the US House of Representatives. We match the speech emphasis with district preferences on key bills using data from the Cooperative Congressional Election Study. We find that House members who rise in opposition to a bill give more passionate speeches when public preferences are aligned with their vote. The results suggest that political actors are not only mindful of public opinion in what they say but also in how they say it.

Keywords: Legislative speech; speech delivery; voter preferences; video-as-data; computer vision

Introduction

Political speech is enormously varied. While political speech is often tedious and boring, some speeches are enthralling and memorable. Whether a speech is dull or captivating depends in no small part on its delivery. Yet, even though the delivery is key in determining how a speech resonates with the public, we lack an understanding of when and why political actors make passionate appeals. In an effort to help fill this gap, this paper focuses on the non-verbal characteristics of political speech, which have rarely been the subject of systematic examination. This gap is due to measurement challenges that researchers face when analyzing non-verbal speech characteristics. Key facets of the delivery cannot be studied using the written record but can only be gauged from video footage. While political scientists have developed a rich toolbox for analyzing the textual features of political speech, the field lacks tools for capturing the non-verbal characteristics of political speech.

To overcome this limitation, we adopt methodological innovations from the field of computer vision for analyzing video data on a large scale. Building on these innovations, we develop and apply an automated video analysis model to analyze video recordings of political speeches. Specifically, we train a convolutional neural network (CNN) to analyze video footage from the US House of Representatives. The CNN is trained to detect gesturing, facial expressions, and pitch to gauge the emphasis in legislative speech.

To understand variation in speech delivery, we develop the argument that legislators adopt particular delivery styles to signal to their constituents. Scholars have long viewed political speeches as signaling tools (Mayhew 1974). Yet, the utility of such signals depends on whether they reach their intended audience. As most political speeches go unnoticed, we argue that legislators rely on emphatic appeals to make their signaling efforts more visible. Particularly in the current media environment, it is imperative that legislators deliver good soundbites to make it past the media gatekeepers or to go viral on social media (Esser, 2008; Larsson, 2020; Negrine and Lilleker, 2002; Strömbäck, 2008). Legislators are aware of this bottleneck and they make fiery appeals when they want to signal their positions. Whether legislators try to send such a signal depends on public opinion. When public opinion is aligned with their policy stance, we expect legislators to make an effort to highlight their position. Thus, in line with existing work on the responsiveness of political speech to public opinion (Bäck and Debus 2018; Baumann et al. 2015; Hill and Hurley 2002), we argue that legislators are not only mindful of public opinion in what they say but also in how they say it.

To assess the effect of public opinion on non-verbal speech characteristics, we link the speech emphasis with survey data. We use multilevel regression and post-stratification, as well as Bayesian additive regression trees and post-stratification to estimate district preferences on a series of bills from the 111th to the 115th Congress (2009–18) using data from the Cooperative Congressional Election Study (Bisbee 2019; Warshaw and Rodden 2012). The results provide support for the notion that legislators employ emphatic appeals to signal their policy positions when they are aligned with constituency opinion.

The findings have important implications for our understanding of legislative speech. Our study is one of few contributions that consider the non-verbal characteristics of legislative speech. In addition to showing that non-verbal speech characteristics contain valuable cues for political research, we highlight how these characteristics are shaped by strategic considerations. Among others, these findings are relevant for researchers trying to gauge the substance of political conflict from speeches (Lauderdale and Herzog 2016; Monroe et al. 2008; Proksch and Slapin 2009). Incorporating the non-verbal characteristics into these efforts can generate novel insights as speech emphasis can help distinguish key policy statements from everyday speech.

Moving Beyond the Textual Features of Political Speech

Political speech is an area of intense research. Particularly the digitization of parliamentary records has helped expand our understanding of the use (Maltzman and Sigelman 1996; Morris 2001; Proksch and Slapin 2012) and substance of parliamentary speech (Hill and Hurley 2002; Morris 2001; Quinn et al. 2010). Despite the undeniable value of these efforts, they are subject to limitations. Efforts to categorize political speech have almost exclusively focused on textual features. While the text of speeches is sufficient for many research questions, political speech has important dimensions that are difficult to capture based on the textual features alone. Key among the characteristics that are typically disregarded in the analysis of speech is the delivery. Speeches are not generally given for the written record. They are a form of political communication where the delivery is central to their intent and effects. While some of the non-textual features may shine through in the written record, much will be lost in the transcription. For example, comparing the written record with video recordings of speeches in the Canadian House of Commons, Cochrane et al. (2022) show that emotional arousal cannot be extracted from the text alone. Succinctly put, researchers have learned a lot about what legislators say but little about how they say it.

Based on the idea that delivery matters for understanding political speech, some contributions have attempted to quantify the non-textual aspects of political speech (Banning and Coleman 2009; Bucy 2016; Wasike 2019) and how they shape perceptions of the speaker (Burgoon et al. 1990; Koppensteiner and Grammer 2010; Masters and Sullivan 1989). These efforts have been

https://doi.org/10.1017/S0007123425100872 Published online by Cambridge University Press

constrained by the difficulty and labor-intensity of manually coding speech recordings. One promising way forward for this research is to build on the recent advances for the automated analysis of audio and video data and to apply these innovations to the ever more widely available digitized recordings of legislative speech.

A nascent literature has begun to employ these tools for researching the non-verbal characteristics of legislative speech and other recordings of political interest. While there are several studies focusing on audio recordings (Dietrich et al. 2019; Dietrich et al. 2019; Knox and Lucas 2021; Rittmann 2024), applications studying video recordings are few and far between. For example, Dietrich (2021) uses video recordings from the US House of Representatives to analyze political polarization. Studying plenary shots, Dietrich finds that legislators have become less likely to mingle across party lines on the House floor as polarization has gone up. Boussalis and Coan (2021) and Boussalis et al. (2021) use computer vision to extract facial expressions of candidates during televised election debates in the United States and Germany, and find that candidates' emotive displays affect viewers' evaluations. Finally, Neumann et al. (2022) analyze politicians' body language in televised ads in the United States, showing that men use more assertive gestures than women.

While existing studies constitute valuable efforts to move beyond the textual features of political speech, the current research agenda using audio and video data is fairly narrow. Due to the novelty of the data and tools for studying digitized audio and video data, the research is heavily invested in validation efforts and in exploring descriptive relationships between actor characteristics and non-verbal political behavior. What is lacking are systematic efforts to situate the new measures in conventional research programs. Indeed, the fact that previous contributions have found substantial variation in non-verbal communication underscores the need for research aimed at explaining variation in the non-verbal characteristics of legislative speech. To this end, we develop and test a theoretical account for emphatic legislative speech.

Signaling Through Emphatic Legislative Speech

The starting point for our theoretical account of emphatic legislative speech is that we conceive of speeches as signals. The signaling function of legislative speech is well established in research on political communication (see, for example, Grimmer 2013; Highton and Rocca 2005; Hill and Hurley 2002), yet there is a notable gap in existing accounts of speech signaling. While there is little doubt that legislators are mindful of public preferences when speaking in public, it has often remained unacknowledged how unlikely it is that such signals will reach their intended audience.

The motivating idea for this contribution is that legislators are aware that the vast majority of speech signals will go unnoticed by the public, which is why legislators can and do make efforts to make their contributions more visible. Particularly in the current media environment, legislators can rely on a push strategy by posting their speeches on their websites or on social media. Legislators can also pursue a pull strategy by trying to create 'broadcastable' moments, hoping that their contributions get picked up by traditional or social media. In practice, push and pull strategies will go hand in hand in that legislators try to create good soundbites, which they then post on social media in the hopes that their messages might go viral.

In trying to create good soundbites, legislators can rely on a host of rhetorical strategies that have been elaborated since antiquity. While captivating oratory can never be separated from the substance of what is being said, it is never a mere textual matter either. Instead, there are a great number of non-verbal characteristics that distinguish a good speech from a bad one. This might cover physical aspects such as gesturing, pose, and body movement, as well as auditory features such as pace, pitch, and volume.

We expect that legislators deliberately rely on these techniques to improve the odds that their signal reaches its intended audience. Notably, we are not interested in whether a signal is actually perceived but whether legislators predictably vary their speech delivery, suggesting a deliberate use

https://doi.org/10.1017/S0007123425100872 Published online by Cambridge University Press

of these techniques for strategic ends. In this sense, we conceptualize speech delivery as a deliberate effort on the part of legislators rather than an unconscious indicator of legislators' true beliefs. To be sure, conceptualizing speech delivery as deliberate should not be equated with insincere. It is easily conceivable that legislators often feel strongly about an issue, resulting in a forceful delivery. Yet, the notion of deliberate emphasis suggests that, for the most part, professional politicians can choose to hide their true feelings when giving a speech if they consider doing so politically advantageous. With regard to our theoretical interest, then, we assume that legislators can choose to send a signal by way of an emphatic delivery.

There is tentative evidence that emphatic and emotional appeals are in fact more likely to be perceived by the public. Given the difficulty of systematically quantifying the emphasis in political speech (Cochrane et al. 2022), there is little direct evidence on this question. The most direct study linking speech emphasis and media visibility is presented by Dietrich et al. (2018). Analyzing audio data from floor speeches in the US House, the authors show that more emotionally charged speeches are more likely to be broadcast and receive media coverage. Beyond the direct evidence, there are various studies on public engagement with political messages, which consistently show that greater emotional intensity predicts the success of political messages on social media (Brady et al. 2017; Heiss et al. 2019; Nave et al. 2018; Peeters et al. 2023), as well as showing that legislators' rhetorical skills and emotional appeals predict their visibility in traditional media (Amsalem et al. 2017; Lupacheva and Mölder 2024; Maier and Nai 2020; Sheafer 2001 2008; Sheafer and Wolfsfeld 2004; Wolfsfeld and Sheafer 2006).

For a first test of our new measure of legislative speech emphasis, we rely on one of the most well-established context factors in research on legislative behavior – the effect of public preferences. We expect that legislators only choose an emphatic delivery when their preferences are aligned with the preferences of their electorate. Only under conditions of alignment should we expect legislators go out of their way to try to send a signal about where they stand politically.[1]

In formulating this expectation, we are able to add nuance to research on legislative signaling. Arguably the most politically consequential signal that legislators can send is their plenary vote. While there is ample evidence that public preferences shape legislators' voting record, legislators often find themselves voting against their district, either because of strongly held personal beliefs or, more commonly, because of pressures from the party leadership, where the pressure to fall in line with the party should be especially strong in the current climate of political polarization. Consequently, while legislators may often find themselves voting against their district, there is little reason to expect legislators to advertise that fact to their electorate by delivering an emphatic speech.

Even though there is little reason to expect legislators to emphatically highlight a vote against the district majority, one might wonder whether legislators with unpopular positions would not be better off not taking the floor at all. While trying to fly under the radar is not an unreasonable strategy, it can also prove dangerous when an unpopular vote comes to light. Therefore, legislators who find themselves voting against their district may prefer to take the floor to explain their vote. At the same time, it would rarely be wise to call attention to an unpopular stance by creating a good soundbite. We can therefore summarize that legislators whose vote is aligned with the preferences of their district are more likely to give an emphatic appeal than legislators with an unpopular stance. As an empirical matter, this expectation also helps distinguish between a deliberate and an unconscious account of speech delivery. If public preferences systematically shape legislators' speech delivery, this is unlikely to result from legislators' true beliefs that accidentally shine through in their delivery.

[1]It should be stressed that whether emphatic legislative speech is more visible to the public is not a precondition for the theoretical expectation to hold. It is enough for legislators to try to signal their position to the public through good soundbites when their preferences align with those of their constituents, irrespective of whether such signaling efforts are ultimately successful.

https://doi.org/10.1017/S0007123425100872 Published online by Cambridge University Press

Table 1. Summary statistics on the bills

					Emphasis	
Bill	Term	Title	Speakers	House vote	Mean	s.d.
HR 1	111th	Recovery and Reinvestment	86	246–183	0.33	0.60
HR 2	111th	State Children's Health Insurance	63	290–135	0.26	0.67
HR 2454	111th	Clean Energy and Security	113	219–212	−0.10	0.61
HR 3590	111th	Comprehensive Health Care Reform	91	219–212	0.15	0.60
HR 4173	111th	Financial Reform	81	223–202	0.19	0.82
HR 2965	111th	End Don't Ask Don't Tell	31	250–175	0.72	0.69
H CR 34	112th	House Budget of 2011	102	235–193	0.59	0.70
HR 2	112th	Repeal Affordable Care	231	245–189	0.39	0.69
HR 6079	112th	Repeal Affordable Care	148	244–185	0.19	0.67
HR 1938	112th	Keystone Pipeline	34	279–147	0.27	0.72
HR 1797	113th	Abortion Bill	22	228–196	0.19	0.60
HR 45	113th	Repeal Affordable Care Act	71	229–195	0.31	0.69
HR 5682	113th	Keystone Pipeline	17	252–161	0.03	0.52
HR 596	114th	Repeal Affordable Care	52	239–186	0.22	0.63
HR 3762	114th	Repeal Affordable Care	50	240–181	0.38	0.56
S 1	114th	Keystone Pipeline	19	270–152	0.29	0.50
HR 36	114th	Pain-Capable Unborn Children Protection	34	242–184	0.15	0.55
HR 3662	114th	Iran Sanctions Act	13	246–181	−0.03	0.42
HR 4760	115th	Securing America's Future	14	193–231	0.43	1.12
HR 36	115th	Pain-Capable Unborn Children Protection	43	237–189	0.06	0.63
HR 1628	115th	American Health Care	119	217–213	0.47	0.69
HR 10	115th	Financial CHOICE	67	233–186	0.17	0.73
HR 3004	115th	Kate's Law	16	257–167	−0.11	0.78
HR 3003	115th	No Sanctuary for Criminals	22	228–195	0.34	0.66
HR 1	115th	Tax Cuts and Jobs Act	91	227–203	0.56	0.73

Note: Speakers refers to the number of speakers during all debates on a bill. House vote presents the result of the final vote on the bill. Mean provides the average emphasis scores across all speeches on a bill, s.d. is the associated standard deviation.

Research Design

Are legislators more likely to deliver emphatic speeches when district preferences on a bill align with their vote? To test this proposition, we study debates on twenty-five pieces of legislation in the 111th–115th US House of Representatives (2009–18). The sample was selected using four criteria. Bills were selected if they were politically salient, if public opinion data on the bills was available, if there was partisan conflict, and if public opinion towards the bill varied between congressional districts. To select the sample, we compiled a list of survey items in the Cooperative Congressional Election Survey (CCES) between 2010 and 2018 where respondents were asked to indicate their preferences on specific pieces of legislation. We then matched these questions to bills in the House of Representatives. These bills overlap to a large extent with votes that were classified as 'key votes' by *Congressional Quarterly* and cover a wide range of domestic and foreign policy issues (Ansolabehere and Jones 2010). From this sample, we discarded bills that were passed without partisan conflict[2] and bills without variation of district-level opinion. The restriction to partisan votes is plausible as voters are more likely to hold or be able to form preferences on important and controversial issues. For the same reason, signaling is a more promising strategy on important and controversial issues, as speeches on irrelevant or undisputed bills are unlikely to be observed by the public. As a practical matter, more speeches are delivered on important and partisan bills.[3] Table 1 lists the twenty-five bills in the resulting sample.

[2]We classify votes as non-partisan if the majority of Republican and Democratic legislators voted for or against a bill, or if more than 30 per cent of legislators did not vote with the majority of their party.

[3]Floor access is comparatively unrestricted in the US House, both in terms of being granted speaking time and in terms of substance (Taylor, 2021). Legislative debate is generally structured by the House Committee on Rules, which specifies the rules under which a bill is brought to the floor. The Rules Committee also determines the length of the debate, which is split equally

https://doi.org/10.1017/S0007123425100872 Published online by Cambridge University Press

In the remainder of this section, we first introduce the dependent variable, the emphasis in legislative speech, and how it can be automatically gauged from plenary video recordings. Next, we discuss the estimation of district preferences on the bills as the independent variable.

Measuring Emphasis in Legislative Speech Using Automated Video Analysis

To study the emphasis in legislative speech, we analyze video recordings of key debates in the House of Representatives. The video recordings of the debates were compiled from the online archives of the House. The sample contains video recordings of all debates on the twenty-five pieces of legislation. We manually discard irrelevant sequences to ensure that we only analyze footage where the camera fully captures the speaker.[4] This results in seventy-seven hours of video footage comprising 2,341 speeches by 543 legislators. Table A2 in the Online Appendix lists the debates and the pieces of legislation.

We employ computer vision to measure the speech emphasis. Specifically, we generalize manual annotations based on a set of training videos to all videos in the sample. We start by drawing an additional sample of 245 speeches by 116 legislators on 37 bills from the 115th Congress as training/test data. We manually selected 245 speeches rather than choosing them at random to ensure sufficient variation in emphasis across our training and test data. This approach allowed us to incorporate a mix of high-, mid-, and low-emphasis speeches in both datasets. The resulting videos were split into 184 training and 61 test videos. Four trained coders were tasked with annotating the emphasis in the speeches using a seven-point Likert scale, ranging from -3 (low emphasis) to $+3$ (high emphasis), for every non-overlapping two-second segment.[5]

Every video was annotated by two randomly selected coders to better judge the emphasis in the videos and to evaluate inter-rater agreement. As continuous annotation of video data is subject to different reaction times and mental processing speeds, annotations can move out of sync. Therefore, we align the annotation sequences by the two coders using the mean absolute error distance. This alignment shifts values by at most two seconds, that is, by one segment. Table 2 summarizes the key values of the manually annotated dataset. On average, the two annotators deviate by less than 0.5 scale points based on the mean absolute error across all two-second segments in the training and test data. Additionally, we report Lin's concordance correlation coefficient and Pearson's correlation coefficient as common measures for inter-rater agreement.[6]

Predicting speech emphasis using a convolutional neural network

To estimate emphasis scores for speeches outside the manually annotated set, we use the training data to train a multimodal convolutional neural network using audio and video inputs. The goal of the network is to assign emphasis scores for each two-second segment. As context information is useful for predicting the current emphasis state of a speaker, we include the surrounding two-second segments for the prediction. Thus, the model takes an input of six seconds of audio and

between proponents and opponents of a bill. The debate is co-ordinated by floor managers, usually the chair and the ranking member of the committee that reported on the bill. Therefore, while the Speaker of the House formally recognizes the individual speakers, floor access is governed by the floor managers, who allocate time to members wishing to address the assembly (Gelman and Goplerud, 2021).

[4]Rules for cutting the videos are documented in Online Appendix A.

[5]Coders watched full speeches and provided annotations every two seconds, such that annotations were contextually informed rather than based on isolated two-second excerpts. The coding scheme for the manual annotations is documented in Table A1 in the Online Appendix.

[6]With the exception of one debate (H.R. 4760, 115th Congress), the annotated speeches are independent of those in the analysis, meaning that the training and test material does not overlap with the analysis data. For those speeches that appear in both, we relied on the human annotations in the analysis. The results remain unchanged if we exclude the speeches on H.R. 4760.

https://doi.org/10.1017/S0007123425100872 Published online by Cambridge University Press

Table 2. Summary metrics for the annotated data set and model evaluation

	Inter-rater baselines		Naïve baselines			
	Train set	Test set	Random guessing*	Zero guessing	Model prediction	30-second aggregate
Number of videos	184	61				
Number of annotated segments	12,686	3,720				
Mean absolute error	0.438	0.438	1.188 ± 0.012	0.932	0.552	0.460
Lin's concordance coefficient	0.816	0.816	− 0.000 ± 0.016	–	0.764	0.837
Pearson's correlation coefficient	0.816	0.818	− 0.000 ± 0.016	–	0.770	0.860

Note: predictions drawn from a clipped standard normal distribution, 1,000 runs.

video data for each two-second segment prediction. Before feeding the data into the network, we perform a series of preprocessing steps which we describe in Online Appendix C.

To predict the speech emphasis, we employ a convolutional neural network.[7] CNNs typically comprise two stages: feature learning and prediction. In the first stage, feature learning, the convolutional base of the model learns hierarchies of modular patterns in the input data. These features are represented in feature vectors that constitute the output of the convolutional base. In the second stage, prediction, this feature vector is fed into a second neural network, which uses the features from the first stage to predict outcome values, in our case, emphasis scores. If trained on a large enough dataset, features learned in the convolutional base are sufficiently generic to be useful for a wide variety of classification tasks. Therefore, especially in cases with small training data, pre-trained networks are commonly used for feature extraction and have proven to be highly effective (Carreira and Zisserman 2017; Chollet and Allaire 2018).

As our manual training data is limited, we use a pre-trained network to extract features for the video input. Specifically, we use a state-of-the art pseudo-3D-Resnet CNN (Qiu et al. 2017). This network is pre-trained on the Kinetics data set, which is commonly used for human action recognition (Kay et al. 2017). The network takes 299 × 299 pixel images as input and generates a 2,048-dimensional feature vector. For the audio input, we use a Soundnet-like subnetwork (Aytar et al. 2016). This network takes the audio input and generates a 512-dimensional feature vector.

After passing the video and audio inputs through these two networks in the feature learning stage, we obtain two feature vectors that summarize the video and audio inputs. In the next step, we combine both vectors and pass them to two fully connected layers. The final layer produces values in the $[-1, +1]$ range. To match the output to the original emphasis scale, we multiply the predicted values by 3 to obtain scores ranging from -3 to $+3$. Figure 1 visualizes the network architecture.

To train the model, we use a mean absolute error loss function. This function minimizes the distance between values predicted by the model and the mean emphasis scores provided by the human annotators. To prevent the neural network from overfitting, we add dropout to the fully connected layers in the second stage (Srivastava et al. 2014).[8] We use the Adam algorithm to optimize the model's parameters (Kingma and Ba 2014).

Applying the model to footage of key vote debates

We apply the trained model to the video footage from all debates in our sample. The network predicts emphasis scores for each two-second sequence. For example, a one-minute speech contains thirty consecutive emphasis scores. To generate one emphasis score per speech, it is

[7]See Torres and Cantú (2022) for an introduction of CNNs in the social sciences.

[8]Dropout is an effective and widely used technique to prevent neural networks from overfitting. Essentially, dropout means randomly setting a number of output features of a layer in a neural network to zero during the training stage. The idea is to add noise to the output values to prevent the network from picking up on patterns that are unique to the training data.

https://doi.org/10.1017/S0007123425100872 Published online by Cambridge University Press

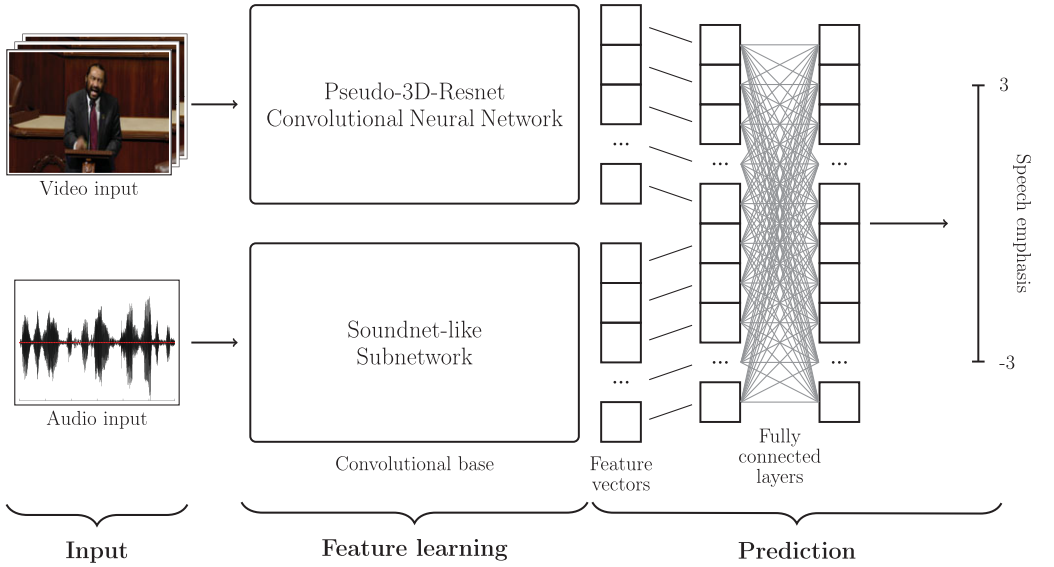

Figure 1. Convolutional neural network architecture.

necessary to aggregate the individual scores. The simplest approach would be to calculate the average emphasis of each speech. Such an approach would ignore that speeches differ in length. This means that a multi-minute speech with thirty seconds of intense delivery would score lower than a one-minute speech with the same sequence. In line with the argument that legislators attempt to signal their issue positions by giving passionate speeches in the hopes of being amplified by traditional or social media, it is sensible to focus on shorter sequences within speeches. Legislators are aware that only short sequences of their speeches may be picked up and broadcast to the public. Therefore, it is sufficient to deliver a short but high-intensity appeal as part of a longer speech, such that short and long speeches with the same high-intensity sequence should score the same. Consequently, we select the thirty-second sequence with the highest within-speech average emphasis to score the speeches.[9] For the same reason, if a legislator delivered more than one speech on a bill, we select the speech with the highest emphasis score for the analysis. We explain this aggregation procedure in more detail in Online Appendix D.

Table 1 provides summary statistics for the resulting data. The number of legislators who delivered speeches on a bill ranges from 13 to 231. Mean emphasis scores range from -0.11 (Kate's Law) to 0.72 (End Don't Ask Don't Tell Act). Figure 2 provides additional information on the distribution of the emphasis scores across all debates, which range from -1.5 to 2.2. Thus, the distribution does not reach the extremes of the emphasis scale, running from -3 to $+3$. This is unsurprising as we average over 30-second segments. The distribution can be characterized as approximately normal with a mean of 0.29 and a standard deviation of 0.70.

Model evaluation

We now turn to the evaluation of the neural network. We present the results of two validation exercises. First, we apply the trained model to the held-out test set of sixty-one videos and compare the model predictions with the human annotations throughout all two-second segments within those speeches. In addition, we apply our aggregation algorithm to all of these speeches

[9]As the cut-off of 30 seconds is arbitrary, we ran additional analyses where we calculate the emphasis scores for 10, 20, 40, 50, and 60-second sequences, as well as the overall average emphasis. The different choices have no effect on the substantive conclusions.

https://doi.org/10.1017/S0007123425100872 Published online by Cambridge University Press

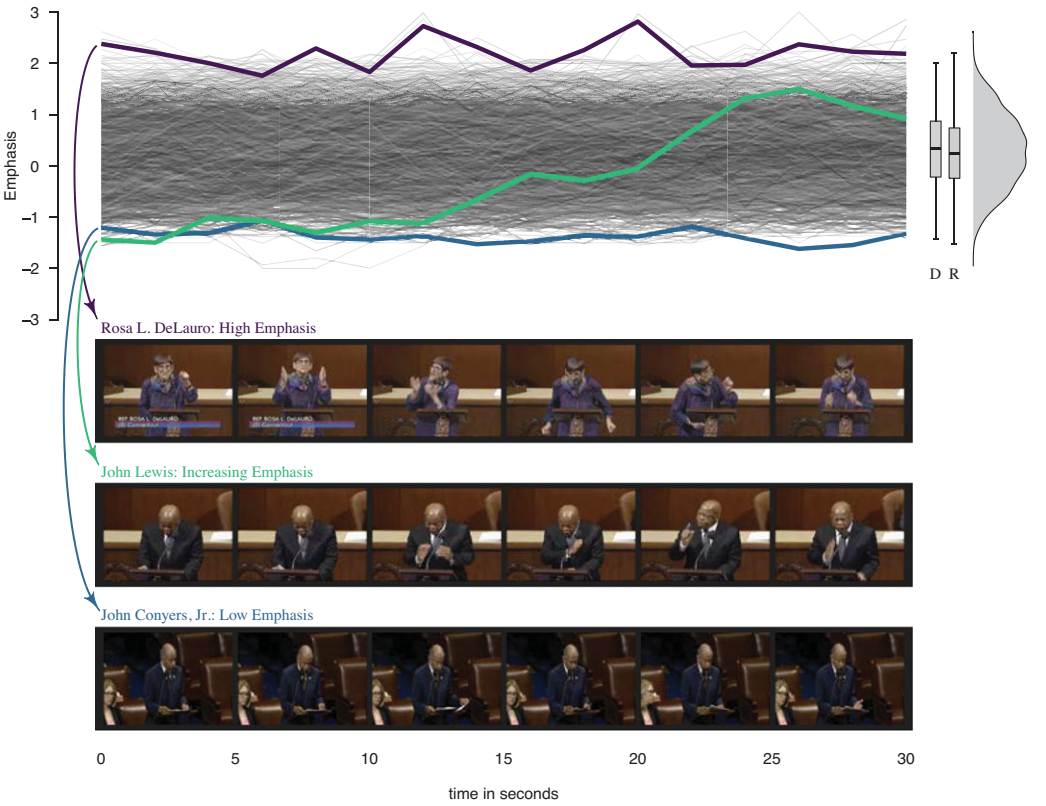

Figure 2. Visualization of the emphasis scores and their distribution.
Note: Each line in the upper panel depicts the estimated speech emphasis over the course of the thirty-second sequences. The three highlighted sequences depict the emphasis scores of the speeches with the highest and lowest average emphasis scores (by Rosa L. DeLauro and John Conyers, Jr.) and the speech with the highest within-speech variance (by John Lewis). The video frames give an impression of how increased levels of gesturing and facial expression are linked to higher estimated emphasis scores. The density curve on the right depicts the distribution of the average emphasis scores as used in the analysis. The two boxplots represent the distributions of average emphasis scores by Democrats (D) and Republicans (R).

using both the model predictions and the human annotations and compare the results. Table 2 shows the results of this comparison, along with the results based on random guessing (drawing values from a clipped standard normal distribution) and zero guessing (predicting an emphasis score of 0). As evaluation criteria, we compute mean absolute errors (MAE), Lin's concordance correlation coefficient (CCC), and Pearson's correlation coefficient (PCC).

Unsurprisingly, the correlations are essentially zero under random guessing. For zero guessing, the correlation is defined as zero as a constant cannot correlate with a variable. For both random and zero guessing, we observe MAE values close to one standard deviation of the underlying label distribution. The neural network achieves considerably lower MAE values. Based on the MAE metric, the machine prediction is 0.552 scale points off from the human annotation when considering two-second segments. This figure is close to the human interrater MAE. For thirty-second aggregates, the MAE decreases further to 0.460, suggesting that aggregation helps average out random noise in the data. Unlike the guessed values, the model predictions show high correlations for both the CCC and the PCC metric. As before, the correlation between the machine prediction and the human annotators is in the same range as the correlation between the human annotators. We thus conclude that the neural network reliably predicts the speech emphasis.

https://doi.org/10.1017/S0007123425100872 Published online by Cambridge University Press

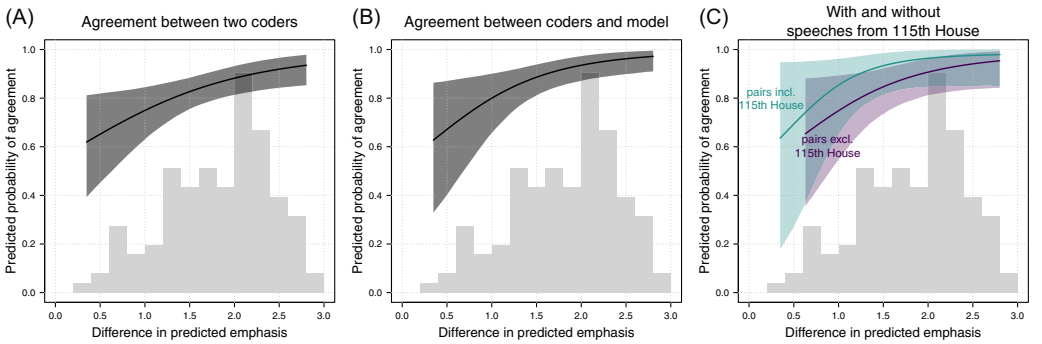

Figure 3. Model evaluation based on pairwise comparisons.
Note: Predicted probabilities and confidence intervals are based on bivariate logistic models, regressing agreement on the difference in predicted emphasis between two speeches. Panel A shows the predicted probability that the two coders agree on which speaker displays greater emphasis. Agreement increases as the model predicts less similar emphasis levels between the two speeches. Panel B is based on pairs where both coders agree on the more emphatic speech, displaying the predicted probability that their ratings align with the model predictions. Agreement between coders and the model increases as the model identifies greater differences in emphasis between the speeches. Panel C distinguishes between pairs that include at least one speech from the 115th legislative term and those that do not. Disagreement between the model and coder ratings is more likely for pairs without speeches from the 115th legislative term.

Our second validation is based on coders' ratings of thirty- rather than two-second segments. We generated 150 speech pairs based on stratified samples from all thirty-second sequences we used in the analysis.[10] For each pair, we asked two coders to indicate which of the speakers displayed a higher level of emphasis, and compared their ratings with the model ratings. Coder and model ratings are considered aligned when the speech identified as more emphatic by the coder receives a higher emphasis score from the model. If the model assigns nearly identical scores to both speeches, we would expect coder ratings to align with the model predictions in around 50 per cent of the cases. As the difference in model emphasis scores increases, we expect coder ratings to be more likely to align with the model predictions.

Naturally, the two coders do not always agree on which of two speeches is more emphatic. Overall, they disagree on twenty-three of the 150 speech pairs. Panel A of Figure 3 plots the probability of agreement between the coders against the difference of predicted emphasis scores by the model. It shows that the probability of the two coders agreeing on the emphasis ordering of a speech pair increases as the speeches are rated more distinct by the model.

Panel B focuses on speech pairs where both coders agree and compares their ratings to the model predictions. As expected, coder and model ratings almost always align when the model predicts a substantial difference in emphasis between the two speeches. However, when the model predicts smaller differences in emphasis, coder ratings are more likely to diverge from the model predictions.

Panel C addresses the fact that the model is trained on speeches from the 115th legislative term, while the analysis includes speeches from the 111th to 115th term. These terms include speakers who were not in the training data, and speaking styles may have evolved, making it more difficult for the model to assess speeches from before the 115th term. To assess whether this is indeed the case, Panel C differentiates between speech pairs that include at least one speech from the 115th legislative term, and that are based on speeches from earlier terms. Indeed, the probability of alignment between coder and model ratings decreases slightly for pairs based on speeches prior to

[10]Stratification is based on the predicted emphasis scores. Fifty pairs consist of random draws from above and below the median of predicted speech emphasis. Another fifty pairs are drawn from the top and bottom 25 per cent of the emphasis distribution, and the remaining fifty pairs come from the top and bottom 10 per cent of the predicted emphasis distribution.

https://doi.org/10.1017/S0007123425100872 Published online by Cambridge University Press

the 115th House. To address this, we include a dummy variable in our analyses, indicating whether a speech is from the 115th term or before.

While the model demonstrates satisfactory performance in the validation exercises, it is not without error. This raises the question of when and why the model makes incorrect predictions. To explore this, we conducted additional quantitative and qualitative analyses of the variation in prediction errors across speeches in our test dataset. We summarize the key insights from this analysis below and provide further details in Appendix E.

First, the model rarely predicts values above $+2$ or below -2, suggesting it may be subject to some attenuation bias. Second, our assessment of the speeches with the highest average prediction error suggests that, compared to human annotators, the model is less sensitive to hand movements occurring in front of the body of the speaker and to gestures made with a closed rather than an open hand. In contrast, the model is more responsive to large, clearly visible hand movements. This observation is plausible as open hands are better visible than closed hands, and gestures stand out more clearly against the background when positioned beside the body of the speaker. Third, we observed that the model tends to overestimate the emphasis of speakers who naturally speak with a strong voice. This is understandable, as a naturally strong voice can resemble an emphatic one. The tendency may correlate with the gender of the speaker, which we account for by controlling for gender in the analysis.

Estimating District-Level Bill Preferences

The key independent variable is the extent to which legislators' votes align with constituency preferences. We draw on survey data from multiple waves of the CCES to estimate district-level preferences. Each survey contains multiple questions on specific bills. Respondents are provided with the title and a short summary of the bill and are asked how they would have voted.[11] Matching each bill in our sample to a CCES item enables us to estimate district preferences towards the bills.[12]

Despite the large sample size of the CCES, it is not designed to be representative of congressional district populations. Thus, simply disaggregating survey answers to estimate district-level preferences would likely yield biased estimates. To overcome this challenge, we rely on the widely used multilevel regression and post-stratification (MrP) for estimating district preferences (Gelman and Little 1997; Lax and Phillips 2009; Warshaw and Rodden 2012). To assess the reliability of the estimates, we complement the MrP approach with Bayesian regression trees and post-stratification (BARP) (Bisbee 2019). We employ MrP and BARP to estimate district-level preferences for the twenty-five bills in our sample. The estimates range from zero to one, where high values indicate high levels of support.[13]

In the next step, we match these estimates to the representatives' voting records on the twenty-five bills. Substantively, we are interested in the extent to which legislators' votes align with the preferences of their electorate. To compute an alignment score, we code roll call votes as one for legislators who voted in favor of a bill and zero for those who voted against it. To assess the extent to which legislators' votes align with the preferences of their electorate, we calculate the absolute difference between legislators' votes (YES $= 1$, NO $= 0$) and the preferences of their districts and subtract this from the maximum distance, 1. We label this variable Vote–District Alignment. Values close to 1 indicate high levels of agreement between legislators and their districts, values close to 0 indicate low levels of agreement. Formally:

[11]The question wording is: 'Congress considered many important bills over the past few years. For each of the following, tell us whether you support or oppose the legislation in principle'.

[12]Table A2 in the Online Appendix reports the matches between the CCES items and the debates in our sample.

[13]Details on the estimation of district preferences are provided in Online Appendix G.

https://doi.org/10.1017/S0007123425100872 Published online by Cambridge University Press

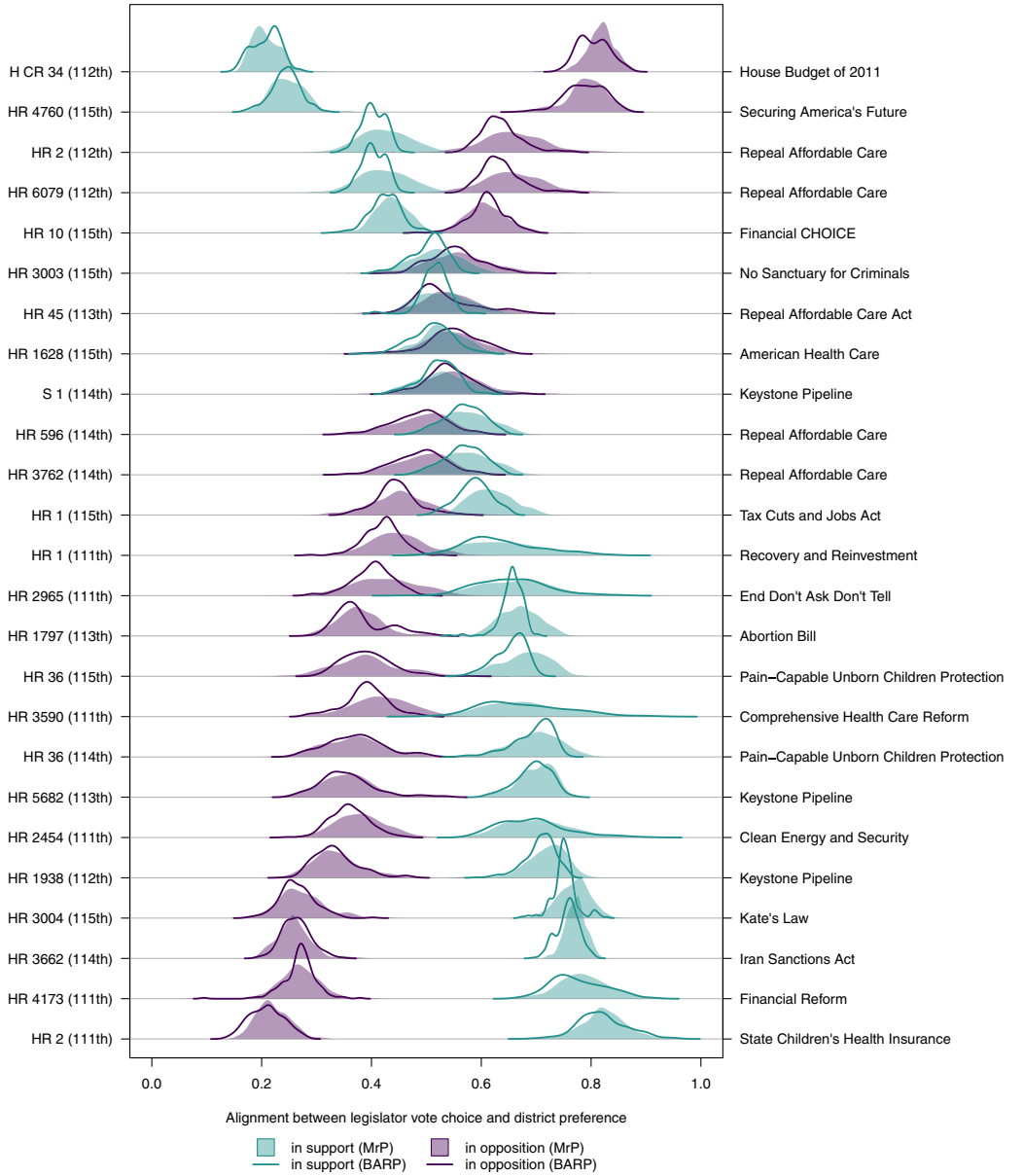

Figure 4. Distributions of alignment between legislator votes and district preferences.

$$\text{VOTE} - \text{DISTRICT ALIGNMENT} = 1 - |\text{YES VOTE} - \text{DISTRICT PREFERENCE}| \quad (1)$$

We present the distributions of the VOTE–DISTRICT ALIGNMENT variable by bill and vote choice in Figure 4. The bright density curves depict the distributions of the VOTE–DISTRICT ALIGNMENT for legislators who voted yes, the dark densities depict the distributions for legislators who voted no. For most bills, MrP and BARP result in similar distributions of the alignment between legislator vote and district preference. In most instances, the distribution for legislators who vote yes diverge from those who vote no. Consider the State Children's Health Insurance Act (HR 2, 111th) as an example. Although there is variation between the districts, almost all districts were fairly supportive of the bill. Thus, the VOTE–DISTRICT ALIGNMENT scores for legislators who voted

https://doi.org/10.1017/S0007123425100872 Published online by Cambridge University Press

for the bill are significantly higher than the values for legislators who voted against the bill. This also means that we observe numerous instances where legislators' votes were misaligned with their districts. Arguably, this finding can be traced back in no small part to the high levels of polarization and the resulting pressure to vote along party lines, such that legislators often find themselves between a rock and a hard place, where they either have to vote against the preferences of their constituents or risk upsetting the party leadership.

District Preferences and Signaling in Legislative Speech

To estimate the effect of district preferences on the emphasis in legislative speech, we proceed in two steps. First, we establish the link between district preferences and speech emphasis by presenting evidence from multilevel models. Next, we increase the complexity of the statistical model to explore the underlying mechanism. Specifically, we investigate whether the association is driven by differences between the delivery styles of legislators from different districts, or because legislators vary their delivery depending on the extent to which their vote is aligned with district preferences. In Online Appendix H.2, we additionally show that our results hold when using a binary measure of vote–district alignment.

We begin by fitting several multilevel models to estimate the effect of district opinion on speech emphasis. The dependent variable is legislators' speech emphasis. The independent variable of interest is the vote–district alignment, indicating the extent to which legislators' floor votes are aligned with district preferences. Following the theoretical argument, we expect vote–district alignment to be positively related to speech emphasis. The model incorporates varying intercepts at the debate level to account for the hierarchical structure of the data, where speeches are nested in legislative debates on the different bills.[14]

To adjust for potential confounding, we consider six control variables. We account for legislators' party affiliation with an indicator variable for Republican legislators. As legislators who vote against the party line may face pressures not to signal this behavior, we include a binary variable that indicates whether a legislator's vote is in line with the majority of their party. We control for seniority to account for legislators' experience in delivering speeches. To account for the possibility that ideologically extreme members might deliver more emphatic speeches than moderate legislators, we include the absolute values of legislators' DW-Nominate scores (Lewis et al. 2020). Finally, we control for gender to account for the possibility that male and female legislators differ in their presentational styles, and a variable indicating whether the speech was held in the 115th term to account for potential differences in measurement error. Table 3 presents eight model specifications. Models (1) to (4) depict the results for speeches in opposition to a bill. Models (1), (3), (5), and (7) are based on MrP estimates, while models (2), (4), (6), and (8) are based on BARP estimates. Models (1), (2), (5), and (6) are baseline models without covariate adjustment. Models (3), (4), (7), and (8) include control variables.

The results indicate that legislators alter their delivery style in reaction to public opinion, but only when they rise in opposition to a bill. Models (1) to (4) show a positive relation between vote–district alignment on legislators' speech emphasis, suggesting that opposing legislators deliver more emphatic speeches when their electorate is also opposed to the bill. While the magnitude of the effect remains relatively stable between the models, the substantive interpretation of the effect is not straightforward. Consider two legislators who both rise in opposition to a bill. In legislator A's district, 65 per cent of the voters are, like legislator A, opposed to the bill. This amounts to a vote-alignment score of 0.65. In contrast, 65 per cent of the voters in legislator B's district support

[14]Observations can also be clustered in legislators when legislators weigh in on multiple debates. We will address this concern in the second part of the analysis.

https://doi.org/10.1017/S0007123425100872 Published online by Cambridge University Press

Table 3. Multilevel specifications with debate random effects. Parentheses report heteroskedasticity consistent wild bootstrap standard errors (Modugno and Giannerini, 2015; Loy, et al. 2023)

	In opposition				In support			
	(1) MrP	(2) BARP	(3) MrP	(4) BARP	(5) MrP	(6) BARP	(7) MrP	(8) BARP
Intercept	− 0.17 (0.14)	− 0.11 (0.14)	− 0.18 (0.24)	− 0.15 (0.23)	0.21 (0.19)	0.27 (0.19)	− 0.77 (0.33)	− 0.70 (0.33)
Vote–district alignment	1.07 (0.25)	0.97 (0.26)	1.30 (0.24)	1.21 (0.28)	− 0.01 (0.29)	− 0.12 (0.30)	0.12 (0.32)	− 0.01 (0.32)
Controls	✗	✗	✓	✓	✗	✗	✓	✓
AIC	1592.75	1596.98	1618.53	1623.06	1716.78	1716.48	1726.01	1726.08
BIC	1611.24	1615.47	1664.74	1669.27	1735.76	1735.46	1773.47	1773.54
Log likelihood	− 792.38	− 794.49	− 799.26	− 801.53	− 854.39	− 854.24	− 853.00	− 853.04
N	751	751	751	751	851	851	851	851
N(debates)	25	25	25	25	25	25	25	25
Var: debates (intercept)	0.04	0.04	0.04	0.04	0.02	0.02	0.03	0.03
Var: residual	0.46	0.47	0.46	0.47	0.42	0.42	0.41	0.41

Note: The dependent variable is the level of emphasis of a legislative speech. AIC: Akaike Information Criterion; BIC: Bayesian Information Criterion.

the bill, which means that only 35 per cent are aligned with their vote,[15] leading us to expect that legislator A delivers a more emphatic speech than legislator B. Based on the estimates from model (3), we expect legislator A to deliver a speech that scores 0.39 (s.e. = 0.07) points higher on the emphasis scale compared to legislator B.[16] This is equivalent to about 0.5 standard deviations of the distribution of speech emphasis among legislators who rise in opposition.

Turning to models (5) to (8) in Table 3, the results suggest that this finding does not generalize to legislators who deliver speeches in support of a bill. The estimated coefficients indicate that when legislators rise in support, they do not deliver their speech more or less emphatically depending on how strongly their district supports the bill. The finding that speeches in support of a bill are less affected by public opinion aligns well with recent research that has highlighted that opposition legislators are more prone to using emotional language (Gennaro and Ash 2022).

Before proceeding with the analysis, we should caution against interpreting the association between district preferences and speech delivery as causal. Establishing causality would require strong theoretical assumptions, especially the absence of confounding variables – assumptions we cannot confidently make given the observational nature of our study. The next section demonstrates that the association holds when considering only within-legislator variation, ensuring that the result is not driven by time-invariant confounders. While this partially addresses concerns about causality, it does not fully resolve them.

Individual Versus Macro-Level Explanation

Having shown evidence for a link between district opinion and speech emphasis among legislators who rise in opposition, we now proceed to test whether the proposed individual-level explanation is driving the findings. It is conceivable that legislators whose preferences are more consistently aligned with those of their constituents might be more likely to signal this alignment to their constituents by giving more emphatic soundbites overall. We are thus interested in differentiating between such an explanation at the legislator level from an explanation where legislators are more emphatic when their position is aligned with their district in a particular debate.

[15]The 0.30 point difference of vote alignment between the two legislators amounts to about two standard deviations of the empirical distribution of the vote-alignment variable.

[16]The standard error is based on simulated first differences for a female Democrat who voted in line with her party before the 115th term. Seniority and ideology are held constant at their mean values (ideology = 0.44, seniority = 15.7).

https://doi.org/10.1017/S0007123425100872 Published online by Cambridge University Press

We adjust the statistical model to study the isolated effects of both explanations. To that end, we partition the total variation of vote–district alignment into two parts: variation *within* districts and variation *between* districts. *Within* vote–district alignment is defined as the deviation of the vote-alignment on a specific bill from legislators' average vote-alignment, which reflects the 'legislator-in-debate' level explanation. *Between* vote–district alignment is defined as legislators' average vote-alignment, which reflects the generic legislator level explanation.

We specify a within–between random effects (REWB) model (Bell et al. 2019; Bell and Jones 2015) to estimate the effects of both variance components in the same model. The model is specified as follows:

$$y_{it} = \beta_0 + \beta_{1W}(x_{it} - \bar{x}_i) + \beta_{2B}\bar{x}_i + \sum_{j=1}^{k} \gamma_j z_{ji} + (\upsilon_i + \upsilon_t + \epsilon_{it}) \quad (2)$$

where y_{it} represents legislator i's emphasis in debate t on a specific bill. x_{it} is the alignment between legislator i's vote after debate t and the preference of their district. $\bar{x}_i$) is the average alignment between legislator i's votes and the preference of their district. υ_i are random intercepts for legislator i and υ_t are random intercepts for debate t. z_{ji} represent the same k individual-level control variables as before.

The model has several properties worth noting. Importantly, the independent variable of interest, VOTE-ALIGNMENT, enters the model in two forms: First, in its de-meaned form $(x_{it} - \bar{x}_i)$, that is, the deviation of the vote-alignment from legislators' average vote-alignment. The coefficient β_{1W} represents the average *within* effect of vote-alignment, that is, the expected change in a legislator's speech emphasis caused by variation of preferences within their district. Thus, positive values of β_{1W} would constitute evidence for the individual-level explanation, making β_{1W} the main coefficient of interest. The coefficient is equivalent to individual fixed effects and is independent of differences in legislators' delivery styles. Second, the model incorporates legislators' average vote-alignment $(\bar{x}_i)$ as a covariate. The coefficient β_{2B} represents the average *between* effect of vote-alignment. This effect captures differences in emphasis between legislators with varying average levels of vote-alignment, that is, legislators with more or less average district support for their floor votes. Thus, positive values of β_{2B} would constitute evidence for the macro-level explanation.

The results in Table 4 provide support for both the individual and the macro-level mechanism. Starting with the individual mechanism, models (1) and (2) show that there is a positive *within* effect of vote–district alignment on speech emphasis. This result indicates that legislators who rise in opposition vary their delivery depending on how closely their vote is aligned with the preference of their district. Specifically, legislators who rise in opposition deliver more emphatic speeches as their districts become increasingly hostile to the bill. Figure 5 helps to assess the size of this effect. Consider a legislator who rises in opposition in two debates on different bills. In the first debate, 50 per cent of the voters in their district are – like them – opposed to the bill. In the second debate, 75 per cent are opposed, making their vote more aligned with public opinion in their district.[17] Based on the estimates from model (1), we would expect the legislator to deliver a speech that scores 0.19 points higher on the emphasis scale during the second debate compared to a speech during the first debate. This is equivalent to 0.45 standard deviations of the de-meaned speech emphasis.

The results also provide evidence for the macro-level mechanism. Both models show a positive *between* effect of vote–district alignment on speech emphasis (1.28 and 1.16 depending on the public opinion estimate). This indicates that legislators whose opposing vote constantly shows high alignment with their electorate tend to deliver more emphatic speeches compared to legislators with less support for their votes. To assess the substantive meaning of this effect,

[17]This 25 percentage point difference is equivalent to about 2.5 standard deviations of the de-meaned vote-alignment variable $((x_{it} - \bar{x}_i))$.

https://doi.org/10.1017/S0007123425100872 Published online by Cambridge University Press

Table 4. Within–between multilevel specifications with legislator and debate random effects. Parentheses report heteroskedasticity consistent wild bootstrap standard errors (Modugno and Giannerini, 2015; Loy, et al. 2023)

	In opposition		In support	
	(1) MrP	(2) BARP	(3) MrP	(4) BARP
Intercept	− 0.29 (0.36)	− 0.26 (0.39)	− 0.51 (0.31)	− 0.45 (0.33)
Vote-district alignment, *within*	0.76 (0.29)	0.77 (0.30)	− 0.22 (0.24)	− 0.27 (0.24)
Vote-district alignment, *between*	1.28 (0.37)	1.16 (0.41)	0.06 (0.27)	− 0.04 (0.29)
Controls	✓	✓	✓	✓
N (Legislators)	295	295	402	402
N (Debates)	25	25	25	25
Var: legislators (intercept)	0.21	0.21	0.18	0.18
Var: debates (intercept)	0.04	0.04	0.02	0.02
Var: residual	0.25	0.25	0.24	0.24
AIC	1456.37	1458.26	1625.39	1625.18
BIC	1511.82	1513.72	1682.35	1682.14
Log likelihood	− 716.18	− 717.13	− 800.69	− 800.59
N	751	751	851	851

Note: The dependent variable is the level of emphasis of a legislative speech. Parentheses show wild bootstrap standard errors.

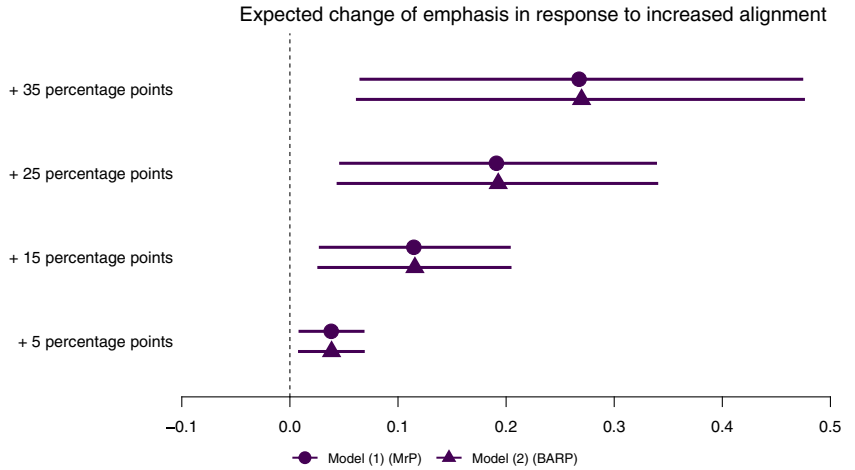

Figure 5. First differences and 95 per cent confidence intervals illustrating the expected change of speech emphasis in response to increased alignment between a legislator's No vote and public opinion in the district.
Note: Wild cluster bootstrap confidence intervals based on model (1) and model (2) in Table 4. The baseline value of the de-meaned district alignment is set to − 0.28 (minimum for the MrP estimates), the mean level of vote-alignment is set to the empirical mean (0.53 for MrP, 0.51 for BARP), Republican is set to zero, vote with party is set to 1, seniority is set to its mean (15.7), gender is set to zero, ideology is set to the empirical mean (0.44).

consider two otherwise similar legislators whose voters differ in their attitudes towards key pieces of legislation, where legislator A enjoys high alignment between her votes and public opinion and B does not. Suppose that the difference in vote alignment between legislators A and B amounts to 25 percentage points on average.[18] Model (1) predicts that this difference has implications for how legislators A and B present themselves on the floor: On average, legislator A would deliver speeches that score 0.32 points higher on the emphasis scale compared to legislator B. This amounts to 0.55 standard deviations of the mean emphasis score.

[18]This is equivalent to about two standard deviations of the legislators' average vote-alignment ($\bar{x}_i$).

https://doi.org/10.1017/S0007123425100872 Published online by Cambridge University Press

Taken together, the results from the REWB model on legislators who rise in opposition provide evidence for both the individual-level and the macro-level mechanism. Regarding the individual-level mechanism, legislators deliver more emphatic speeches as their vote becomes more aligned with their district. At the same time, legislators whose votes constantly show high alignment with their districts deliver more emphatic speeches on average.

The results from models (3) and (4) echo the findings from the models in Table 3. They show that this finding does not generalize to legislators who rise in support of a bill. The magnitudes of the estimated within and between estimates are not distinguishable from zero at conventional levels of statistical significance.

Conclusion

Automated analyses of audio and video data have begun to make their way into political science research (Dietrich 2021; Dietrich et al. 2019; Dietrich et al. 2019; Knox and Lucas 2021,Nyhuis et al. 2021). These techniques promise to bring about significant innovations in a number of research fields by allowing scholars to make better use of the massive amounts of digitized data and to move beyond the narrow focus on digitized political text. In research on legislative speech specifically, incorporating the new tools and data sources enables the systematic study of questions beyond the substance of speeches and a greater appreciation of the non-verbal aspects of political speech.

In this paper, we have built on these nascent efforts to explain variation in the delivery of legislative speech. We have argued that legislators are not only strategic in what they say, but also in how they say it. As legislators are aware that most speeches go all but unnoticed, they make conscious decisions about when to deliver emphatic speeches in order to increase their chances of being featured in the media. Constituency preferences are a key factor in explaining such signaling in legislative speech. As actors with a singular interest in re-election, legislators are only expected to highlight their positions when they align with the preferences of their constituents.

To assess whether the speech delivery is responsive to public opinion, we relied on automated video analyses to measure the extent to which legislators deliver emphatic speeches on twenty-five key bills in the 111th–115th US House of Representatives (2009–18). The analyses have shown consistent effects of constituency opinion on speech delivery. Across different model specifications, legislators rising in opposition to a bill were found to deliver more emphatic speeches, the more their districts are opposed to the measure.

Having shown evidence for the effect of district preferences on the non-verbal characteristics of legislative speech in the US House, one might wonder to what extent the effect generalizes to other legislatures. We would argue that the effect of constituency preferences on legislative speech should be strongest where legislators can expect to benefit the most from a personal vote (cf. Carey and Shugart 1995), such that the US House probably constitutes a most likely case for observing an effect of public preferences on speech signaling. To a somewhat lesser extent, one might expect that legislators are more mindful of their messaging in legislatures where rhetoric is more valued, such that assemblies such as the UK House of Commons is probably characterized by more emphatic appeals than its continental European counterparts (Yildirim 2025; Osnabrügge et al. 2021).

Despite consistent effects across different model specifications, a few limitations should be explicitly addressed. First, we argued that the Cooperative Congressional Election Study is useful for estimating district preferences on Congressional roll call votes and that attitudes towards specific bills are better suited for gauging constituency preferences and their effects on legislative speech than a general ideology measure. It should not be left unmentioned, however, that using these indicators comes at a price. As the CCES only features survey items on key congressional votes, our analysis is restricted to key debates, raising the question whether our findings generalize

https://doi.org/10.1017/S0007123425100872 Published online by Cambridge University Press

beyond debates on important bills. On the one hand, there is little reason to expect strategic legislators to go out of their way to signal their position when it clashes with the preferences of their constituents. On the other hand, speeches on inconsequential bills might generally be characterized by fewer emphatic appeals, which could result in fewer differences between speeches of legislators who agree with their constituents and those who do not. Future research could shed light on the question of whether our findings generalize beyond key bills by building on the present efforts and studying a broader sample of bills, while relying on a coarser measure of district ideology. Such research is greatly simplified by the promises of computer vision where trained neural networks can easily be deployed to study speeches on other bills.

Second, the observational nature of our study prevents us from making causal claims about the relationship between district preferences and emphatic speech delivery. This limitation is shared by many studies on the effect of public opinion on elite behavior, as public opinion can rarely be experimentally manipulated. However, future research might identify scenarios where naturally occurring exogenous variation in public opinion offers more credible support for the assumptions required to establish causal links between public opinion and speech delivery (cf. Hager and Hilbig 2020). Relatedly, our research design does not offer insights into how accurately legislators assess public opinion in their district before giving a speech, leaving unanswered questions about the precise mechanism underlying our findings.

Third, future research should also try and link the textual and the non-verbal characteristics of legislative speech more closely in order to gain additional insights into legislative speech. While our study has made first steps towards such an analysis by showing how the non-verbal characteristics are tied to position-taking in speeches, additional research could investigate which specific parts of speeches legislators choose to emphasize and which content features betray a high-energy delivery.

Finally, while our study is among the first to examine the determinants of emphatic legislative speech, it does not distinguish between kinesic and vocalic cues in shaping emphasis. Since kinesic cues are conveyed through video and vocalic cues through audio, future research on non-verbal components of political speech might benefit from disentangling their relative contributions. In particular, methodological investigations into how audio and video modalities influence model performance could provide valuable insights into the distinct informational content carried by gestures and vocal inflection. Such analyses could help researchers prioritize modalities when computational or analytical constraints necessitate focusing on just one modality.

Overall, the study of non-verbal characteristics with emerging computer vision tools holds enormous promise for research on political speech, legislative behavior, and more. The present contribution constitutes one of the first attempts to systematically trace and explain the non-verbal characteristics of legislative speech. In line with previous research, our findings underscore that legislators are conscious and strategic in their use of legislative speech and that such strategy is not exhaustively described by the substantive aspects of legislative speech. To further refine our understanding of speech delivery, we hope that the theoretical and empirical advances presented in this contribution elicit a growing interest in the analysis of non-verbal aspects of political speech.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0007123425100872.

Data Availability statement. Replication data for this article can be found in Harvard Dataverse at: https://doi.org/10.7910/DVN/5EWYTN.

Acknowledgments. We thank Morten Harmening, Felix Münchow, Marie-Lou Sohnius, Sam Känner, and Caro Krömer for excellent research assistance. We are grateful to Thomas Gschwend, Lukas Stoetzer, Christian Arnold, Seo-young Silvia Kim, Patrick Kraft, Marcel Neunhoeffer, Anna Adendorf, Oke Bahnsen, Sean Carey, Franziska Quoß, Viktoriia Semenova, Christoph Steinert, and the participants of the Severyns Ravenholt Seminar in Comparative Politics at the University of

Washington, as well as the participants of the American Politics Research Group at the University of North Carolina at Chapel Hill for feedback on earlier versions of this manuscript.

Financial support. This project was supported by the Deutsche Forschungsgemeinschaft and the Sonderforschungsbereich 884. Further support is gratefully acknowledged from the research alliance ForDigital.

Competing interests. None.

References

Amsalem E, Sheafer T, Walgrave S and Loewen PJ (2017) Media motivation and elite rhetoric in comparative perspective. *Political Communication* **34**, 385–403.

Ansolabehere S and Jones PE (2010) Constituents' responses to congressional roll-call voting. *American Journal of Political Science* **54**, 583–597.

Aytar Y, Vondrick C and Torralba A (2016) Soundnet: Learning sound representations from unlabeled video. *Advances in Neural Information Processing Systems*, 892–900.

Banning S and Coleman R (2009) Louder than words: A content analysis of presidential candidates' televised nonverbal communication. *Visual Communication Quarterly* **16**, 4–17.

Baumann M, Debus M and Müller J (2015) Convictions and signals in parliamentary speeches: Dáil Éirann debates on abortion in 2001 and 2013. *Irish Political Studies* **30**, 199–219.

Bäck H and Debus M (2018) Representing the region on the floor: Socioeconomic characteristics of electoral districts and legislative speechmaking. *Parliamentary Affairs* **71**, 73–102.

Bell A and Jones K (2015) Explaining fixed effects: Random effects modeling of time-series cross-sectional and panel data. *Political Science Research and Methods* **3**, 133–153.

Bell A, Fairbrother M and Jones K (2019) Fixed and random effects models: Making an informed choice. *Quality & Quantity* **53**, 1051–1074.

Bisbee J (2019) BARP: Improving mister P using Bayesian additive regression trees. *American Political Science Review* **113**, 1060–1065.

Boussalis C and Coan TG (2021) Facing the electorate: Computational approaches to the study of nonverbal communication and voter impression formation. *Political Communication* **38**, 75–97.

Boussalis C, Coan TG, Holman MR and Müller S (2021) Gender, candidate emotional expression, and voter reactions during televised debates. *American Political Science Review* **115**, 1242–1257.

Brady WJ, Wills JA, Jost JT, Tucker JA and Van Bavel JJ (2017) Emotion shapes the diffusion of moralized content in social networks. *Proceedings of The National Academy of Sciences of The United States of America* **114**, 7313–7318.

Bucy EP (2016) The look of losing, then and now: Nixon, Obama, and nonverbal indicators of opportunity lost. *American Behavioral Scientist* **60**, 1772–1798.

Burgoon JK, Birk T and Pfau M (1990) Nonverbal behaviors, persuasion, and credibility. *Human Communication Research* **17**, 140–169.

Carey JM and Shugart MS (1995) Incentives to cultivate a personal vote: A rank ordering of electoral formulas. *Electoral Studies* **14**, 417–439.

Carreira J and Zisserman A (2017) Quo vadis, action recognition? A new model and the kinetics dataset. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6299–6308.

Chollet F and Allaire JJ (2018) *Deep Learning with R*. Manning.

Cochrane C, Rheault L, Godbout JF, Whyte Tanya, Wong MWC and Borwein S (2022) The automatic analysis of emotion in political speech based on transcripts. *Political Communication* **39**, 98–121.

Dietrich BJ (2021) Using motion detection to measure social polarization in the U.S. House of Representatives. *Political Analysis* **29**, 250–259.

Dietrich BJ, Schultz D and Jaquith T (2018) This floor speech will be televised: Understanding the factors that influence when floor speeches appear on cable television. Paper presented at the workshop *"PoliInformatics: New Data. New Methods, New Insights"*.

Dietrich BJ, Hayes M and O'Brien DZ (2019) Pitch perfect: Vocal pitch and the emotional intensity of congressional speech. *American Political Science Review* **113**, 941–962.

Dietrich BJ, Enos RD and Sen M (2019) Emotional arousal predicts voting on the U.S. Supreme Court. *Political Analysis* **27**, 237–243.

Esser F (2008) Dimensions of political news cultures: Sound bite and image bite news in France, Germany, Great Britain, and the United States. *International Journal of Press/Politics* **13**, 401–428.

Gelman A and Little TC (1997) Poststratification into many categories using hierarchical logistic regression. *Survey Methodology* **23**, 127–135.

https://doi.org/10.1017/S0007123425100872 Published online by Cambridge University Press

Gelman D and Goplerud M (2021) United States: Evolving determinants of participation in floor debates. In Bäck H, Debus M and Fernandes JM. (eds.), *The Politics of Legislative Debates*. Oxford: Oxford University Press, pp. 801–824.
Gennaro G and Ash E (2022) Emotion and reason in political language. *The Economic Journal* **132**, 1037–1059.
Grimmer J (2013) *Representational Styles in Congress: What Legislators Say and Why it Matters*. Cambridge.
Hager A and Hilbig H (2020) Does public opinion affect political speech? *American Journal of Political Science* **64**, 921–937.
Heiss R, Schmuck D and Matthes J (2019) What drives interaction in political actors' facebook posts? Profile and content predictors of user engagement and political actors' reactions. *Information, Communication and Society* **22**, 1497–1513.
Highton B and Rocca MS (2005) Beyond the roll-call arena: The determinants of position taking Congress. *Political Research Quarterly* **58**, 303–316.
Hill KQ and Hurley PA (2002) Symbolic speeches in the U.S. Senate and their representational implications. *Journal of Politics* **64**, 219–231.
Kay W, Carreira J, Simonyan K, Zhang B, Hillier C, Vijayanarasimhan S, Viola F, Green T, Back T, Natsev P, Suleyman M and Zisserman A (2017) The kinetics human action video dataset, arXiv preprint arXiv: 1705.06950, 1–22.
Kingma DP and Ba J (2014) Adam: A method for stochastic optimization, arXiv preprint arXiv: 1412.6980, 1–15.
Knox D and Lucas C (2021) A dynamic model of speech for the social sciences. *American Political Science Review* **115**, 649–666.
Koppensteiner M and Grammer K (2010) Motion patterns in political speech and their influence on personality ratings. *Journal of Research in Personality* **44**, 374–379.
Larsson AO (2020) Winning and losing on social media: Comparing viral political posts across platforms. *Convergence* **26**, 639–657.
Lauderdale BE and Herzog A (2016) Measuring political positions from legislative speech. *Political Analysis* **24**, 374–394.
Lax JR and Phillips JH (2009) How should we estimate public opinion in the States? *American Journal of Political Science* **53**, 107–121.
Lewis JB, Poole K, Rosenthal H, Boche A, Rudkin A and Sonnet L (2020) Voteview: Congressional roll-call votes database. https://voteview.com.
Loy A, Steele S and Korobova J (2023) "lmeresampler: Bootstrap methods for nested linear mixed-effects models. *R package version 0.2.4*.
Lupacheva T and Mölder M (2024) A place to speak and be heard? Parliamentary speech and media attention in Estonia, 2011–2019. *Legislative Studies Quarterly* **49**, 905–924.
Maier J and Nai A (2020) Roaring candidates in the spotlight: Campaign negativity, emotions, and media coverage in 107 national elections. *International Journal of Press/Politics* **25**, 576–606.
Maltzman F and Sigelman L (1996) The politics of talk: Unconstrained floor time in the U.S. House of Representatives. *The Journal of Politics* **58**, 819–830.
Masters RD and Sullivan DG (1989) Nonverbal displays and political leadership in France and the United States. *Political Behavior* **11**, 123–156.
Mayhew DR (1974) *Congress: The Electoral Connection*. Yale University Press.
Modugno L and Giannerini S (2015) The wild bootstrap for multilevel models. *Communications in Statistics: Theory and Methods* **44**, 4812–4825.
Monroe BL, Colaresi MP and Quinn KM (2008) Fightin' words: Lexical feature selection and evaluation for identifying the content of political conflict. *Political Analysis* **16**, 372–403.
Morris JS (2001) Reexamining the politics of talk: Partisan rhetoric in the 104th House. *Legislative Studies Quarterly* **26**, 101–121.
Nave NN, Shirman L and Tenenboim-Weinblatt K (2018) Talking it personally: Features of successful political posts on facebook. *Social Media and Society* **4**, 1–12.
Negrine R and Lilleker DG (2002) The professionalization of political communication: Continuities and change in media practices. *European Journal of Communication* **17**, 305–323.
Neumann M, Fowler EF and Ridout TN (2022) Body language and gender stereotypes in campaign video. *Computational Communication Research* **4**, 254–274.
Nyhuis D, Ringwald T, Rittmann O, Gschwend T and Stiefelhagen R (2021) Automated video analysis for social science research, *Handbook of Computational Social Science*, **2**, Routledge pp. 386–398.
Osnabrügge M, Hobolt SB and Rodon T (2021) Playing to the gallery: Emotive rhetoric in parliaments. *American Political Science Review* **115**, 885–899.
Peeters J, Opgenhaffen M, Kreutz T and van Aelst P (2023) Understanding the online relationship between politicians and citizens: A study on the user engagement of politicians' facebook posts in election and routine periods. *Journal of Information Technology and Politics* **20**, 44–59.
Proksch SO and Slapin JB (2009) Position taking in European parliament speeches. *British Journal of Political Science* **40**, 587–611.
Proksch SO and Slapin JB (2012) Institutional foundations of legislative speech. *American Journal of Political Science* **56**, 520–537.

Qiu Z, Yao T and Mei T (2017) Learning spatio-temporal representation with pseudo-3d residual networks. *Proceedings of the IEEE International Conference on Computer Vision*, pp. 5533–5541.

Quinn KM, Monroe BL, Colaresi M, Crespin MH and Radev DR (2010) How to analyze political attention with minimal assumptions and costs. *American Journal of Political Science* **54**, 209–228.

Rittmann O (2024) Legislators' emotional engagement with women's issues: gendered patterns of vocal pitch in the German Bundestag. *British Journal of Political Science* **54**, 937–945.

Rittmann O, Ringwald T and Nyhuis D (2025) Replication Data for: Public Opinion and Emphatic Legislative Speech: Evidence from an Automated Video Analysis. https://doi.org/10.7910/DVN/5EWYTN, Harvard Dataverse, V1.

Sheafer T (2001) Charismatic skill and media legitimacy: An actor-centered approach to understanding the political communication competition. *Communication Research* **28**, 711–736.

Sheafer T (2008) Charismatic communication skill, media legitimacy, and electoral success. *Journal of Political Marketing* **7**, 1–24.

Sheafer T and Wolfsfeld G (2004) Production assets, news opportunities and publicity for legislators: A study of Israeli knesset members. *Legislative Studies Quarterly* **29**, 611–630.

Srivastava N, Hinton G, Krizhevsky A, Sutskever I and Salakhutdinov R (2014) Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research* **15**, 1929–1958.

Strömbäck J (2008) Four phases of mediatization: An analysis of the mediatization of politics. *International Journal of Press/Politics* **13**, 228–246.

Taylor AJ (2021) Legislative Speech in Presidential Systems. In Hanna B, Debus M and Fernandes JM (eds.), *The Politics of Legislative Debates*. Oxford: Oxford University Press, pp. 51–71.

Torres M and Cantú F (2022) Learning to see: Convolutional neural networks for the analysis of social science data. *Political Analysis* **30**, 113–131.

Warshaw C and Rodden J (2012) How should we measure district-level public opinion on individual issues? *The Journal of Politics* **74**, 203–219.

Wasike B (2019) Gender, nonverbal communication, and televised debates: A case study analysis of Clinton and Trump's nonverbal language during the 2016 town hall debate. *International Journal of Communication* **13**, 251–276.

Wolfsfeld G and Sheafer T (2006) Competing actors and the construction of political news: The contest over waves in Israel. *Political Communication* **23**, 333–354.

Yildirim TM (2025) Emotions in the aisles: Unpacking the use of emotive language in the UK House of Commons. *European Journal of Political Research* **64**, 943–959.

Cite this article: Rittmann O, Ringwald T, and Nyhuis D (2025) Public Opinion and Emphatic Legislative Speech: Evidence from an Automated Video Analysis. *British Journal of Political Science* 1–21. https://doi.org/10.1017/S0007123425100872