This Floor Speech Will Be Televised: Understanding the Factors that Influence When Floor Speeches Appear on Cable Television*

Bryce J. Dietrich† Dan Schultz‡ Tracey Jaquith§

Abstract

Members of Congress (MCs) use floor speeches to signal their colleagues and constituents, but we know very little about the way floor speeches reach the American public. Using 152,138 hours of audio from 74,158 floor speeches and 1,974 cable news programs, we show both verbal and non-verbal emotional expressions influence the likelihood floor speeches appear on CNN, Fox News, and MSNBC. More specifically, we find emotionally intense speeches – as indicated by a speaker's vocal pitch – (1) are more likely to be aired, (2) generate more total coverage, (3) receive more air time, and (4) reach a larger audience. In doing so, we introduce original software – called the *AudioNewser* – which allows users to query the *Internet Archive*'s unprecedented television news collection using their own audio files. Ultimately, this helps expand the study of legislative speech to include text *and* audio data.

Word Count: $\sim 9,957$

*Comments and suggestions welcome. We are grateful to Brian Fogarty and conference or seminar participants at Rice University and the PoliInformatics workshop for helpful feedback. We also thank the hardworking men and women at the *Internet Archive*. Without them, this study would not have been possible. In particular, we would like to thank Brewster Kahle (Digital Librarian). We would also like to thank everyone who helped make the *TV News Archive* possible, including Roger Macdonald (Director), Nancy Watzman (Managing Editor), Robin Chin (News Archivist and Researcher), and Katie Dahl (Research Associate).

†Department of Political Science, University of Iowa, 341 Schaeffer Hall, Iowa City, IA 52242 (bryce-dietrich@uiowa.edu, http://www.brycejdietrich.com).

‡Senior Software Engineer, Internet Archive, 300 Funston Avenue, San Francisco, CA 02138 (schultzd@media.mit.edu, https://slifty.com).

§TV Architect, Internet Archive, 300 Funston Avenue, San Francisco, CA 02138.

1 Introduction

Members of Congress (MCs) use floor speeches to signal their colleagues and constituents (Mayhew, 1974), but in order to achieve the latter people must actually view the speeches themselves. Unfortunately, whether it is radio, television, or the internet, we know very little about the way floor speeches reach the American public. Using original software and 145,706 hours of cable news broadcasts, the present study considers what types of floor speeches are aired on CNN, Fox News, and MSNBC. Ultimately, we find emotionally intense speeches – as indicated by a speaker's vocal pitch – generate the most coverage, suggesting verbal *and* non-verbal expressions likely fit within the broader dramatic narratives cable news networks are trying to create.

Floor speeches are not inherently newsworthy. Not only are there literally hundreds of floor speeches each day, but it also is difficult to say which floor speeches best represent the nature of a given legislative debate. In these instances, news organizations are more likely to select floor speeches which elicit an emotional response from viewers (Graber and Dunaway, 2017). Similar to images (Graber, 1996) and music (Grabe et al., 2000), we argue floor speeches that "provoke emotional responses and physiological stimulation or arousal among the members of the audience" are more likely to be aired by CNN, Fox News, and MSNBC (Uribe and Gunter, 2007, 209). Indeed, Graber (1994) found news directors only covered legislative sessions when there were intense emotional exchanges, otherwise floor debates were "not really worth covering" (496).

Ultimately, two types of speeches likely fit within the dramatic narratives news organizations are trying to create (Graber and Dunaway, 2017). First, speeches delivered with more emotional intensity – as indicated by the speaker's increased vocal pitch – are more likely to be aired. Not only is vocal pitch associated with emotional activation (for review, see Mauss and Robinson, 2009), but previous studies have suggested a reporter's tone can produce a response in viewers similar to other dramatic devices like fast-paced editing and

close-ups (Grabe, Zhou, and Barnett, 2001). For example, Burgoon (1978) suggested a "dynamic style," exemplified by high intensity, high pitch, and faster speaking rate, influenced extroversion judgements, as well as perceptions of reporter competence. Ultimately, "these results suggest that voice tone is an important delivery attribute that may affect viewers' attention and arousal" (Grabe, Zhou, and Barnett, 2001, 643).

Second, speeches which express negative sentiment are also more likely to be aired. Even though news organizations undoubtedly want to inform and educate, they also need to turn a profit (Hamilton, 2004; Sparrow, 1999). Negativity is one way to achieve this end as news organizations compete for increasingly smaller and more homogenous audiences (e.g., Cohen, 2008). Not only do viewers tend to find negative news more interesting for a variety of reasons (Galtung and Ruge, 1965), but news organizations also prefer to broadcast conflict and negativity (e.g., Harcup and O'Neill, 2016), especially when both are directed towards Congress (e.g., Robinson and Appel, 1979). Since MCs often employ a similar narrative (Fenno, 1977), we expect CNN, Fox News, and MSNBC are more likely to air floor speeches which cast Congress in a negative light.

Using 6,432 hours of audio from 74,158 floor speeches and every cable news program broadcast between January 6, 2009 and August 4, 2014, we demonstrate speeches delivered with more emotional intensity – suggested by higher vocal pitch – are more likely to be aired. Moreover, speeches which express negative sentiment also tend to appear on cable news broadcasts, with some evidence suggesting negativity towards Congress is especially likely to be aired. In doing so, we introduce original software called the *AudioNewser* (or an audio browser for news) which allows users to query every cable news broadcast using their own audio files. The present paper uses the *AudioNewser* to identify which floor speeches appear on cable news, but our software could also be used to determine whether certain political advertisements, *YouTube* videos, or other audio/video files of interest are also aired.

This paper proceeds as follows. In Section 2, we discuss existing scholarship on con-

gressional news coverage. We use this literature to develop our key theoretical predictions regarding what floor speeches are likely to be aired on cable news. These hypotheses are outlined in Section 3. We explain our novel data in Section 4 which includes the largest audio collection from floor speeches and cable news broadcasts ever compiled. In this section, we also introduce the software we created for this study. Our results can be found in Section 5. Ultimately, we show cable news organizations have strong preferences for certain types of verbal and non-verbal expressions. We conclude in Section 6 with a summary and discussion of avenues of future research. More details about our data, as well as additional results, can be found in the Supplemental Information (SI).

2 Emotional Expression and Congressional Media Coverage

The Value of Covering Congress

Whether from television, radio, newspapers or their own sources, journalists are inundated with information. Part of their job is to select the small amount of information that becomes news. This process – called "gatekeeping" – involves "selecting, writing, editing, positioning, scheduling, repeating and otherwise massaging information to become news" (Reese, Vos, and Shoemaker, 2009, 73). Since the result is a narrow snapshot of the political world, it is vital to understand why certain types of information are selected (Shoemaker and Vos, 2009). Not only do most voters learn about their own representatives from the news (Waismel-Manor and Tsfati, 2011), but news coverage can help frame debates (Sellers, 2000) and generate external constituent pressure (Wawro, 2001). Consequently, understanding what types of floor speeches are aired will help us better understand whether some MCs and the issues they advance have an inherent advantage.

In doing so, we also learn which news values influence the gatekeeping process (Harcup and O'Neill, 2016). For example, economic pressures may entice news organizations to

only cover certain hot-button issues (e.g., Hamilton, 2004). This insistence on "relevancy" (see Harcup and O'Neill, 2016, 1482) would suggest the majority of congressional coverage is determined by the issues of the day, but it can also be driven by the characteristics of the members themselves. Since journalists assume the actions of the "powerful elite" have a greater consequence (see Harcup and O'Neill, 2016, 1471), both seniority (Kuklinski and Sigelman, 1992) and institutional position (Cook, 1989; Johnson and O'Grady, 2013) likely also influence congressional coverage. As politics continues to become an increasingly mediated process (Bennett and Entman, 2001), this too likely gives some MCs and the issues they advance an inherent advantage.

No study has considered whether the gatekeeping process also influences the types of floor speeches that are televised, even though we know some undoubtedly do. Sen. Susan Collins' (R-ME) 44-minute speech in support of Justice Kavanaugh's nomination is one of the most recent examples. Within 4 minutes of when she started speaking on October 5, 2018, Fox News began live streaming the speech on their *YouTube* channel and posted a clip of her speech on their website. They then reposted the video two more times that day. CNN and MSNBC also posted videos of the speech on their respective *YouTube* channels and websites with CNN posting clips seven times online and MSNBC posting Sen. Collins' full speech three times. By the end of the day on October 7th, the speech had been viewed a total of 449,772 times on each network's *YouTube* channel. CNN, Fox News, and MSNBC also showed portions of the speech on over fifteen different programs between October 5th and 6th (see Figure 1), suggesting that floor speeches not only appear on these networks, but when they do they have the potential to reach a large audience.

Even though journalists likely take into consideration a number of factors when deciding what floor speeches to broadcast, we argue verbal and non-verbal emotional expressions also likely influence the gatekeeping process. Not only can emotional expressions be used to underline the importance of floor debates, but such clips are also more likely to "go viral" (Berger and Milkman, 2012; Kim, 2015) which give journalists a strong incentive to select

Figure 1: Cable News Programs Broadcasting Sen. Collins' Speech on October 5-6, 2018

Note: Frames from Sen. Collins' speech that appeared on cable news. Sorted alphabetically using the program title. Five clips were found for each network.

speeches delivered with emotional intensity. Although there are a variety of ways to measure emotional intensity, we use a speaker's vocal pitch which has consistently been shown to be associated with emotional activation (for review, see Mauss and Robinson, 2009). For example, the mean vocal pitch for Sen. Collins' speech was 226.69Hz, but the clips shown in Figure 1 were delivered at 242.37Hz. This suggests CNN, Fox News and MSNBC broadcast clips in which Sen. Collins was speaking with greater emotional intensity, as compared to the rest of her speech.

This is consistent with the way audio and video recordings are increasingly used by news organizations (Harcup and O'Neill, 2016). Although sometimes audio and video recordings can become the story, for the most part, they are often used to *emphasize* "a number of news factors, such as bad news, drama or surprise" (Harcup and O'Neill, 2016, 10). Instead of broadcasting "novel" floor speeches in which MCs are yelling and screaming (for review, see Caple and Bednarek, 2013), most floor speeches that appear on air are similar to the speech delivered by Sen. Collins in which she is clearly emotionally invested in the points she is advancing, but she is not yelling and screaming. Journalists are more likely to select such speeches because MCs seem to genuinely care about the points they are advancing when they speak at a heightened vocal pitch. This makes emotionally intense speech especially useful when trying to *emphasize* the "drama" unfolding during floor debates.

The Value of Verbal and Non-Verbal Emotional Expressions

Additional empirical support is found in several studies which show emotional messages are more likely to be stored in long-term memory (e.g., Bradley et al., 1992; Lang, Dhillon, and Dong, 1995) and "that the arousal level of a message (be it a commercial, a public service announcement, or a news brief) may be as much a determinant in whether it is remembered, as the valence of the appeal" (Lang, Dhillon, and Dong, 1995, 324). Consequently, viewers should find emotionally intense speeches – as indicated by changes in vocal pitch – more engaging and memorable. This ultimately makes emotionally intense speeches more likely

to be selected from the rest of the speeches delivered during an important debate.

Previous studies of the vocal qualities of news broadcasters also suggest a general preference for animated voices (Burgoon, 1978; Martín-Santana, Reinares-Lara, and Reinares-Lara, 2017). For example, Warhurst, McCabe, and Madill (2013) suggest a broadcaster's voice can help "form an image of the person speaking" (5). Here, "the four Ps–Pause, Pitch, Pace and Projection" are essential to making the broadcaster seem more "animated" (5) We argue the same applies to floor debates. Just as news organizations prefer "animated" broadcasters, viewers likely prefer more "animated" floor speeches. Consequently, floor speeches delivered with more emotional intensity – as indicated by increased vocal pitch – are more likely to be aired on CNN, Fox News, and MSNBC.

We also expect negativity to have a similar effect. Not only is negative information often thought to be more "newsworthy" (Soroka, 2012), but it can also better keep the public's attention (e.g., Geer, 2008). Consistent with the "conflict" news value (for review, see Caple and Bednarek, 2013), such presentation styles are more likely to "grab and maintain the attention of those individuals who may be passively 'flipping' from one channel to the next with their remotes" (Forgette and Morris, 2006, 447). Even though this preference for "high-conflict presentation styles" tends to make the majority of news coverage negative (Patterson, 2011), we expect this general trend to be more pronounced when this negativity is directed towards Congress.

Not only do MCs often speak disparagingly of Congress (Fenno, 1977), but the vast majority of congressional coverage tends to be negative (Cappella and Jamieson, 1996). For example, Rozell (1994) found that "the press has generally held Congress in low esteem" and that "over the years press coverage of Congress has moved from healthy skepticism to outright cynicism" (109). Robinson and Appel (1979) and Lichter and Amundson (1994) come to similar conclusions. The former analyzed congressional news coverage during a one-month period in 1976 and found that network television coverage of Congress was much more negative than print coverage. Similarly, Lichter and Amundson (1994) analyzed all

television news stories about Congress from 1972 to 1992 and found that 90 percent of all evaluations of Congress were critical.

We again use Sen. Collins' speech as our motivating example. Even though the clips shown in Figure 1 were delivered with more emotional intensity – as indicated by the increase in vocal pitch – Sen. Collins' remarks were not especially negative. However, this was not the only clip that consistently appeared on CNN, Fox News, and MSNBC shortly after her speech concluded. In another popular clip, Sen. Collins said that the confirmation process had "become so dysfunctional it looks more like a caricature of a gutter-level political campaign than a solemn occasion." This clip – which is generally negative towards Congress – was shown on nine programs between October 5th and 6th, ultimately suggesting verbal *and* non-verbal emotional expressions likely influence what floor speeches are aired.

3 Theoretical Expectations

Although many news values influence the gatekeeping process, we are interested in whether the general preference for negative news, conflict, and emotionality make certain floor speeches more likely to appear on air. Indeed, journalists increasingly incorporate emotionality into their reporting (Pantti, 2010; Wahl-Jorgensen, 2013; Peters, 2011) which is why we expect to find the same when they cover floor debates. Ultimately, we expect certain types of verbal and non-verbal expressions will generate more coverage, even though we fully acknowledge the importance of other variables.

First, non-verbal expressions often trigger comparable emotional responses in viewers (Uribe and Gunter, 2007) which increases the likelihood the story is remembered (e.g., Lang, Dhillon, and Dong, 1995). This gives journalists a strong incentive to broadcast speeches delivered with emotional intensity, especially since such clips are more likely to "go viral" (Berger and Milkman, 2012; Kim, 2015). This leads to our first hypothesis:

H1: Cable news organizations are more likely to broadcast floor speeches deliv-

ered at a higher vocal pitch.

However, news organizations also want to broadcast engaging content which grabs the attention of a shrinking news audience (e.g., Cohen, 2008). Not only are dramatic narratives (Graber and Dunaway, 2017) and negative information (Soroka, 2012) often thought to be more "newsworthy," but the latter is also thought to garner more attention (e.g., Geer, 2008). Consequently, news coverage tends to be negative (Patterson, 2011) which leads to our second hypothesis:

> H2: Cable news organizations are more likely to broadcast floor speeches which express a negative sentiment.

We expect to find this general relationship will be more pronounced when MCs speak negatively about Congress. Not only do "members of Congress run for Congress by running *against* Congress" (Fenno, 1977, 168), but journalists have increasingly shown a preference to cast Congress in a negative light (Cappella and Jamieson, 1996). Consequently, we expect floor speeches that do the same are more likely to be broadcast which leads to our final hypothesis:

> H3: When members of Congress express negativity towards Congress, their floor speeches are *more* likely to be aired by cable news organizations.

To test our hypotheses, we created original software – called the *AudioNewser* – which allows users to query every cable news broadcast since 2009 using their own audio files. We use this software to determine which floor speeches appear on CNN, Fox News, and MSNBC but it can be used more broadly to see whether certain political advertisements, *YouTube* videos, or other audio/video files of interest make similar appearances. Our novel data, methodology, and software are introduced in the next section.

4 Data and Measurement

We collected 6,432 hours of audio from 863 U.S. House debates beginning in January 6, 2009 and ending in August 4, 2014 from *HouseLive*. We restricted our analysis to speeches that had at least 50 words. Although shorter speeches could have some substantive value, they are unlikely to be noticed by CNN, Fox News, and MSNBC. Ultimately, this yielded the text and audio from 74,158 speeches delivered by 619 MCs.

Our 145,706 hours of television data was obtained from the *Internet Archive*. Using a known television schedule, the *Internet Archive* records one-minute segments using dedicated servers based in San Francisco, CA. These segments are then combined into full programs, ultimately yielding every program aired on local and national broadcasts. For this study, we used 1,974 CNN, Fox News, and MSNBC programs broadcasted between January 6, 2009 and August 4, 2014.[1]

The *AudioNewser* – An Audio Browser for News

Similar to *Shazam*, the *AudioNewser* uses an audio hashing (or fingerprinting) algorithm to reduce an audio signal into a series of representative strings. Ultimately, we employ a matching algorithm developed by Cotton and Ellis (2010) which uses the paired locations of spectrogram peaks. These peaks are then incorporated into a matching pursuit (MP) algorithm which iteratively selects localized peaks that correspond with the most energetic points in a signal. This is different than simply selecting spectrogram peaks within a fixed window. Instead, a MP algorithm is able to identify salient features across varying time-frequency scales which are then incorporated into a common searchable database.

A "landmark" is defined using the two center frequencies for the paired spectrogram peaks and the time difference between those centroids. For every 32 block of time steps (which are approximately 1 second) the 15 most pronounced paired spectrogram peaks are selected. These are then systematically compared to each other, ultimately yielding approximately 45

[1] Additional details about our speech and television data can be found on pages S4–S6 in the SI.

landmarks per second. These are then converted into a unique 20-bit hash – with the first 8 bits being reserved for the first frequency, the second 6 bits for the frequency difference, and the final 6 bits reserved for the time difference. All landmarks are stored in a hash table with an identification number associated with each video and a time offset value which is simply the location of the landmark (in seconds) relative to the start of the video.

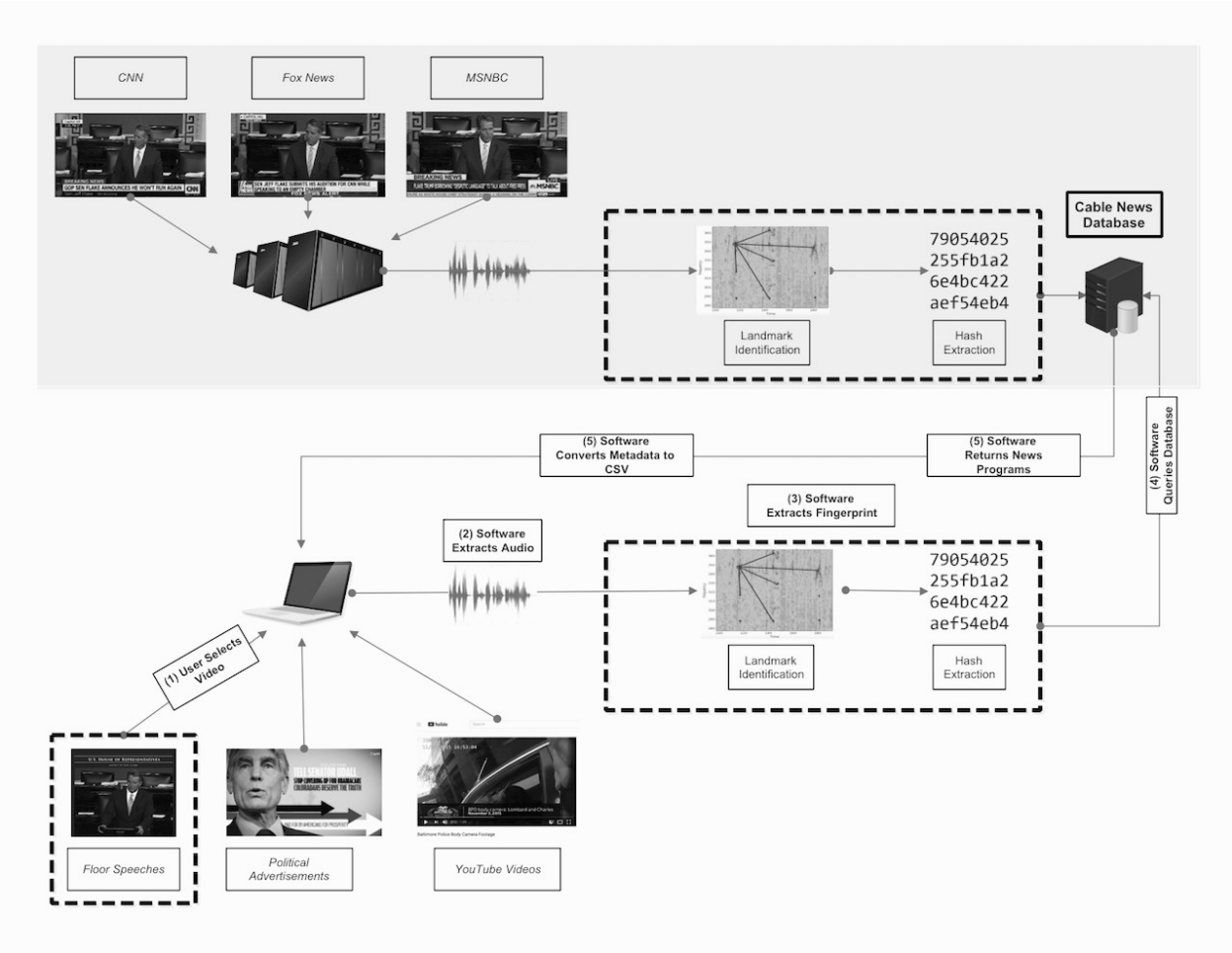
Once the hash table is created it can be queried using a user-provided video. This is done by first decomposing the video using the aforementioned MP algorithm, then dividing the video into five-second (non-overlapping) clips and identifying the relevant landmarks. This means each clip will have an average of 225 landmarks. Those landmarks are then compared to the large hash table built from all television news programs with likely matches returning multiple matched landmarks for each clip. These matches have approximately the same offset times since each are scaled relative to the starting point of each clip, meaning they can be easily searched.

Figure 2 outlines our software. In the section highlighted in grey, we show how we created the cable news database we used for this study. This required recording live cable news streams using multiple servers, extracting the audio, and then applying the Cotton and Ellis (2010) algorithm to create a searchable database composed of millions of hexadecimal strings. Once the *AudioNewser* is installed (see unhighlighted section), the user can query the cable news database by first selecting a video file from their local computer. The software then extracts the audio, applies the Cotton and Ellis (2010) algorithm, and returns the cable news programs in which the user's video appears.[2]

Ultimately, floor speeches were aired 496 times on CNN (59), Fox News (219), and MSNBC (218). Of those, 258 appeared only once and the average floor speech appeared 1.92 times which can be seen in Figure S7 on page S33 of the SI. We are mostly interested in the number of times floor speeches are aired since news organizations frequently broadcast

[2]We also created an additional tool which automatically generates storyboards to help users quickly verify the output. More details about this tool can be found on page S8 of the SI.

Figure 2: Explaining How the *AudioNewser* (or the Audio News Browser) Works



Note: An overview of the software we created for this study. In the section highlighted in grey, we show how the cable news database was created. In the unhighlighted section, we show the basic user experience. For more details, please see pages S5–S8 of the SI.

different clips from the same speech. For example, two clips from Sen. Collins' speech on Justice Kavanaugh were consistently aired on CNN, Fox News, and MSNBC which makes 496 the more relevant count. With that said, the *vast* majority of floor speeches do *not* appear on cable news, but Figure S8 on page S36 of the SI shows those that do reach around 856,796 viewers which gives speeches that are aired a substantial reach.

Vocal Pitch and Emotional Intensity

Our central hypothesis focuses on the emotional intensity of a speech which we measure using a speaker's vocal pitch. Similar to Dietrich, Enos, and Sen (2018), we extracted vocal pitch using *Praat*.[3] This popular open-source speech analysis program estimates the fundamental frequency (commonly known as vocal pitch) by dividing the autocorrelation of a windowed signal by the autocorrelation of the window itself. Dietrich, Enos, and Sen (2018) not only show *Praat* can be consistently used to extract vocal pitch, but it has also been used by others to achieve similar ends (Vogel et al., 2009). For these reasons, *Praat* is by far the most popular speech analysis software on the market which is why we used it for this study.

When individuals become emotionally activated, a typical physiological response is a tightening of the vocal cords. This tightening, in turn, leads to a higher-than-average vocal pitch when speaking which is why one of "the most consistent associations reported in the literature is between arousal and vocal pitch, such that higher levels of arousal have been linked to higher-pitched vocal samples" (Mauss and Robinson, 2009, 222). Like many physiological responses, this tightening of the vocal cords is thought to occur largely below conscious awareness making minor changes in vocal pitch an "inherently honest indicator" of a speaker's "internal state" (Ekman et al., 1991, 133-134). Indeed, "several studies have shown that...the tone of a person's voice leaks information that is not revealed by the verbal content or facial expressions associated with the message" (Zuckerman and Driver, 1985, 129). On pages S19–S29 of the SI, we include an extended discussion of vocal pitch and report the results from several validation exercises. We also provide some descriptive statistics on pages S35–S39.

Similar to Dietrich, Enos, and Sen (2018), in order to control for gender-based differences in vocal pitch we standardized vocal pitch using each MC's mean and standard deviation. For example, imagine Nancy Pelosi (D-CA) gave three speeches, each of which were delivered at:

[3] `http://www.fon.hum.uva.nl/praat/`

175Hz, 200Hz, and 225Hz. Her mean vocal pitch would be 200Hz with a standard deviation of 25Hz. In all the models outlined below we scaled the average vocal pitch of each speech using these baseline measures. For these example speeches, that would yield -1, 0, and 1, suggesting Nancy Pelosi was less emotionally intense in her first speech, whereas the inverse is true for her third speech. Additional details about pitch extraction and the standardization process can be found on pages S5 of the SI.

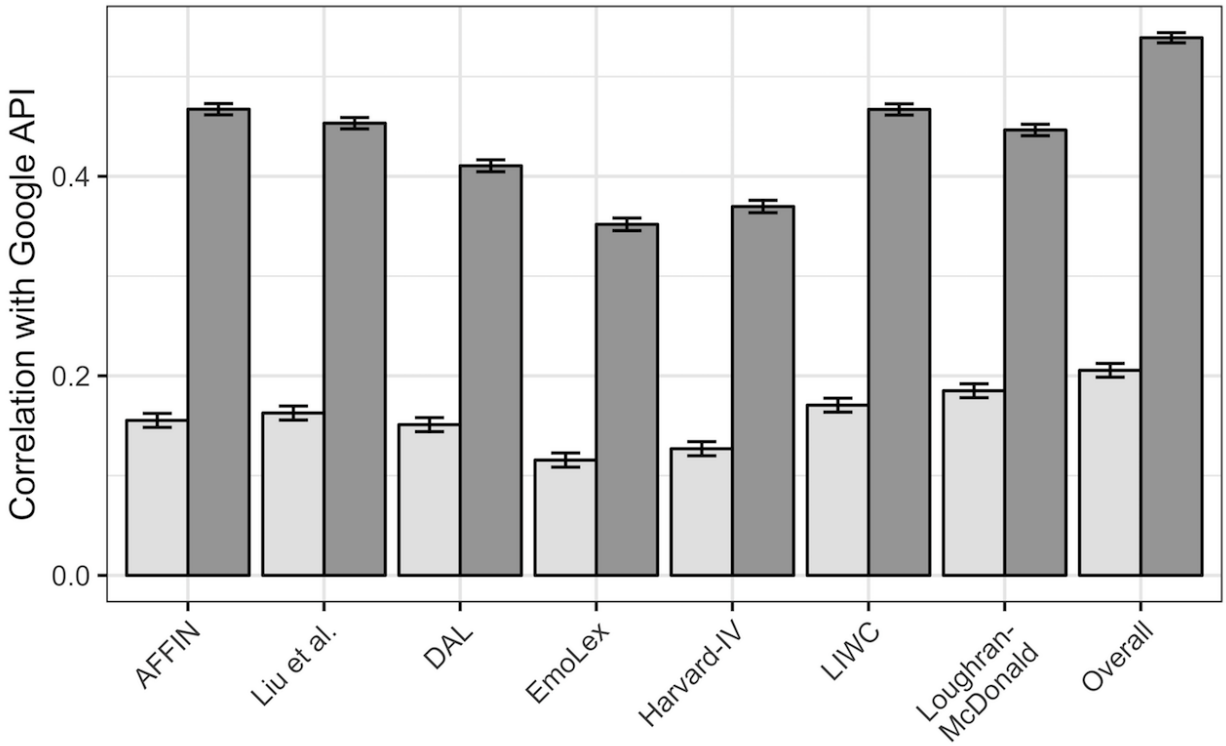
Entity Recognition and Positive/Negative Sentiment

Sentiment towards Congress (see `Congress Sentiment`) was measured using the Google Cloud Natural Language Application Programming Interface (Google API). On pages S8–S18 of the SI, we provide more information about the Google API and the `analyzeEntitySentiment` function which we used for this paper. More specifically, we used this function to obtain the average sentiment expressed towards the following keywords and phrases: "Congress," "United States House of Representatives," "Senate," "United States Senate," and "Capitol Hill." This yielded a score ranging from -1 to 1 with positive values implying the speech was generally more positive towards Congress. The speech's `Overall Sentiment` was determined using the average sentiment expressed towards all named entities referenced in the speech. Here too, the score ranges from -1 to 1 with higher values implying the speaker generally expressed more positive sentiment.

Ultimately, 25,755 speeches had `Overall Sentiment` scores greater than zero, whereas only 17,983 speeches had `Overall Sentiment` scores less than zero, suggesting floor speeches were more likely to be positive. The inverse is true for congressional references. Here, 2,656 speeches had `Congress Sentiment` scores greater than 0 and 6,701 had `Congress Sentiment` scores less than 0, suggesting MCs rarely cast Congress in a positive light. To make these text-based measures more comparable to our vocal pitch measure, we standardized both sentiment scores using each MC's mean and standard deviation with positive values

implying MCs were *more* negative than we would expect.[4]

Figure 3: Comparing the Google API to Seven Sentiment Dictionaries



Note: In the light (see) and dark (see) grey bars we show the correlation between overall and congressional sentiment as measured by the Google API and seven dictionaries. 95 percent confidence intervals are shown. All correlations are statistically significant at the 0.001-level. For more details, see pages S13–S18 of the SI.

We validate our text-based measures in Figure 3. Here, we created overall and congressional sentiment scores using the words from seven dictionaries. To construct the former, we simply applied each dictionary, then took the percentage of positive words and subtracted the percentage of negative words. This yielded a measure that ranged from -1 to 1 with the

[4]For the congressional sentiment scores, zeros were entered when a speech did not make a congressional reference. On pages S52–S55 of the SI, we re-estimate all of our models with no imputation yielding similar results, even though `Congress Sentiment` is no longer statistically significant at the 0.10-level. Since `Vocal Pitch`, `Overall Sentiment`, and `Congress Sentiment` are always included in the same model this equates to only including speeches which mention Congress. This would eliminate 65.43 percent of our data (or 48,520 speeches) which is why we report the imputed results.

latter implying 100 of the words were positive and none of the words were negative. We constructed the congressional sentiment score in a similar way, but instead of using words from the whole document we only used the 10 words surrounding the keywords and phrases outlined above. This also yielded a score which ranged from -1 to 1 with the latter implying 100 percent of the words surrounding congressional references were positive and none of the words were negative. On pages S13–S18 of the SI, we provide examples from each dictionary and a detailed example of how both comparison measures were constructed.

Regardless of the dictionary we used, the results obtained from the Google API were significantly correlated with all dictionary-based measures. For example, when we used the Linguistic Inquiry Word Count (LIWC) dictionary we found the Google API overall sentiment score was significantly correlated with the "positive" and "negative" emotion categories. More specifically, when we took the former and subtracted the latter, we found the result correlated with the overall sentiment derived from the Google API at the 0.001-level ($\rho = 0.47$, $t = 143.82$, df = 74,144, $p < 0.001$). We found essentially the same result when we used the mean score from all seven dictionaries (see "Overall" columns). Here, we scored each speech using all seven dictionaries, then took the average and determined whether it was correlated with the Google API output. Not only was the result again significantly correlated at the 0.001-level, but the correlation increased to 0.54 suggesting the Google API yields reasonable results.

Even though our congressional sentiment score was significantly correlated with our dictionary-based measures, the magnitude is noticeably smaller. Given that, we used Democrats who were in the 111th and 112th Congresses to help further validate our measure of congressional sentiment. We ultimately expect Democrats will be *more* negative towards Congress when they move from the majority party in the 111th Congress to the minority party in the 112th since they lost institutional control. This is exactly what we found. More specifically, the mean congressional sentiment for Democrats decreased 129.83 percent from -0.02 in the 111th Congress to -0.04 in 112th Congress, suggesting they were significantly more

negative when they were in the minority ($t = 5.72$, $df = 6,933$, $p < 0.01$). A similar decline did not happen for Republicans, who actually became 9.14 percent more positive in the 112th Congress (-0.03) as compared to the 111th (-0.04), suggesting their sentiment towards Congress improved when they became the majority party ($t = -0.84$, $df = 5,095$, $p > 0.05$). We think these results again suggest the Google API yields reasonable results.

Additional Controls

To help isolate the effect of a speech's emotional content, we included a dummy variable which returns a 1 for the Speaker of the House, Majority/Minority Leaders and Majority/Minority Whips. Otherwise, this variable returns a zero. A similar dummy variable was created for committee chairs with 1 indicating the MC was the chair of a committee and 0 otherwise. This variable was created using data provided by Stewart and Woon.[5] Previous studies have also found seniority (e.g., Kuklinski and Sigelman, 1992) and ideological extremity (e.g., Vos, 2014) can also drive coverage which is why we included the number of years served and the absolute value of the difference between each MC's DW-Nominate score and their party medians as controls. Seniority was obtained from *GovTrack*, whereas ideology and party identification were obtained from *Voteview*. We included the latter as a control because floor speeches have also been shown to be influenced by partisanship (e.g., Harris, 2005). For similar reasons, we include dummy variables for whether the speech was less than one minute and whether it was delivered during an election year, both of which have been shown to influence speaking behavior (e.g., Maltzman and Sigelman, 1996).

Even though there is considerable debate over whether male MCs receive more news coverage than their female counterparts (Vos, 2014), we also included a dummy variable indicating whether *GovTrack* identified the MC as a male (1) or female (0). A similar control was created for race using the House website since it too has been hypothesized to influence media coverage (e.g., Gershon, 2012). Along these lines, we controlled for the

[5] `http://web.mit.edu/17.251/www/data_page.html`

number of major bills (as measured by CQ Weekly's "Bills to Watch") since we expect cable news organizations will more actively cover the House floor when important bills are being debated. A control was also included for the speech duration (in minutes) since spikes in vocal pitch will have a greater effect on mean vocal pitch when a speech is shorter.

We estimated a Structural Topic Model (STM) to control for a speech's issue content (Roberts, Stewart, and Tingley, 2014). More specifically, we created a variable capturing the extent to which Democratic and Republican legislators were speaking on issues owned by their respective parties. Ultimately, this variable returns a value ranging from 0 to 1 with 1 implying 100 percent of the speech was dedicated to party issues. We include this variable to not only control for the issue content of each speech, but also to control for any partisan bias in network coverage. More details about the STM and the actual topics we used can be found on pages S41–S44 of the SI.

5 Results

We measure television coverage in four ways. These measures are (1) whether a floor speech appeared on television, (2) the number of times the floor speech was televised, (3) the total air time (in minutes) dedicated to the speech, and (4) the estimated number of total viewers (in millions) based on 2014 Nielsen ratings. The first variable was modeled using a rare events logistic regression (called `ReLogit`) in order to account for the dichotomous and unbalanced nature of the dependent variable (King and Zeng, 2001). A negative binomial regression was used to model the second variable.[6] Tobit regressions were used to model the last two variables since both are censored at zero. All models were estimated using Zelig with standard errors clustered around each MC to account for speech-level dependence (Imai, King, and Lau, 2008).[7] Additional details about all of the variables we used in this

[6] On pages S47–S52 of the SI, we also estimated zero-inflated negative binomial regressions. These models yielded similar results.

[7] We chose to use clustered standard errors because there is currently no multilevel implementation for `ReLogit`. On pages S55–S57 of the SI, we re-estimated our models including a randomly varying intercept

study can be found on pages S29–S32 and network specific results can be found on pages S45–S47 of the SI. Our main results are reported in Table 1.

Consistent with our first hypothesis, Models 1.1 (or Table 1, Model 1) and 1.2 (or Table 1, Model 2) show floor speeches delivered with more emotional intensity are significantly ($p < 0.01$) more likely to be aired. Based on our results from Model 1.1, when MCs' vocal pitch is 1 standard deviation above their baseline and all other variables are held constant their speeches are 1.86 times more likely to appear on CNN, Fox News, and MSNBC as compared to floor speeches delivered at their baseline (predicted probabilities of 0.005 when `Vocal Pitch` is set to 1 and 0.003 when `Vocal Pitch` is set to 0). This holds even after accounting for party identification, ideology, institutional position, seniority, race, and whether it was an election year (see Model 1.2).

Even though the probability of a given speech appearing on television is low, vocal pitch seems to increase the likelihood a floor speech is aired. Indeed, the speakers who appeared on CNN, Fox News, and MSNBC spoke on average 0.822 standard deviations *above* their baseline, whereas the speakers who did not appear spoke at 0.003 *below* their baseline. Although this difference is statistically significant at the 0.001-level ($t = 13.30$, $df = 71,196$, $p < 0.001$), the raw magnitude is slight enough that the speakers are unlikely aware of these changes in their vocal pitch. Not only does this provide additional support for our first hypothesis, but these results are also consistent with Models 1.3–1.8 which show floor speeches delivered with more emotional intensity are also (1) more likely to be aired more than once, (2) receive more air time, and (4) reach more viewers.

To help interpret our results, Figure 4 shows predicted values from Table 1, Models 1 (see "Aired'), 3 (see "Coverage"), 5 (see "Minutes"), and 7 (see "Viewers"). For the dark grey bars, we held all other variables constant and calculated predicted values when the speaker

using both `Stata` and `R`. We also re-estimated every model in `Stata` using robust standard errors clustered around each MC. In total, `Vocal Pitch` is a statistically significant ($p < 0.01$) predictor of television coverage in 87 of 89 models which gives us a great deal of confidence our main result is not due to our modeling choices.

Table 1: CNN, Fox News, and MSNBC Coverage of Verbal and Non-Verbal Emotional Expressions on the Floor of the U.S. House of Representatives

	Televised		Total Coverage		Total Minutes		Total Viewers	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	-5.883^{**}	-6.013^{**}	-5.421^{**}	-6.293^{**}	-3.512^{**}	-3.548^{**}	-16.053^{**}	-15.724^{**}
	(0.074)	(0.307)	(0.139)	(0.433)	(0.442)	(0.436)	(1.834)	(2.073)
Vocal Pitch	0.617^{**}	0.605^{**}	0.857^{**}	0.882^{**}	0.289^{**}	0.281^{**}	1.315^{**}	1.272^{**}
	(0.057)	(0.058)	(0.093)	(0.101)	(0.040)	(0.037)	(0.194)	(0.196)
Overall Sentiment	0.237^{**}	0.226^{**}	0.269^{**}	0.318^{**}	0.095^{**}	0.095^{**}	0.497^{**}	0.502^{**}
	(0.061)	(0.058)	(0.090)	(0.106)	(0.033)	(0.033)	(0.145)	(0.141)
Congress Sentiment	$0.181^{\dagger}$	$0.177^{\dagger}$	$0.186^{\dagger}$	0.142	$0.076^{\dagger}$	$0.075^{\dagger}$	$0.314^{\dagger}$	$0.310^{\dagger}$
	(0.102)	(0.102)	(0.106)	(0.115)	(0.042)	(0.045)	(0.170)	(0.181)
Republican		$-0.332^{\dagger}$		$-0.677^{\dagger}$		-0.132		-0.853
		(0.200)		(0.358)		(0.112)		(0.568)
|DW-Nominate|		1.197^{**}		1.745^{**}		0.443^{*}		2.242^{*}
		(0.415)		(0.604)		(0.206)		(1.023)
Party Issue		-0.073		-0.134		-0.030		-0.130
		(0.063)		(0.101)		(0.028)		(0.130)
Seniority		0.022^{**}		$0.016^{\dagger}$		0.010^{**}		0.034^{*}
		(0.006)		(0.008)		(0.004)		(0.016)
House Leader		1.899^{**}		2.090^{**}		0.857^{**}		3.964^{**}
		(0.195)		(0.425)		(0.202)		(0.807)
Committee Chair		0.197		0.252		0.099		0.352
		(0.223)		(0.304)		(0.117)		(0.562)
Male		-0.222		0.204		-0.075		-0.101
		(0.181)		(0.297)		(0.121)		(0.493)
White		-0.195		0.173		-0.065		-0.366
		(0.230)		(0.382)		(0.130)		(0.548)
CQ Bills		0.212^{**}		$0.195^{\dagger}$		0.109^{**}		0.329^{*}
		(0.051)		(0.100)		(0.031)		(0.138)
One Minute		-0.053		0.070		0.025		-0.254
		(0.172)		(0.292)		(0.063)		(0.305)
Duration		-0.116^{*}		-0.160^{*}		-0.028		-0.274^{*}
		(0.048)		(0.080)		(0.019)		(0.114)
Election Year		-1.005^{**}		-1.015^{**}		-0.402^{**}		-2.122^{**}
		(0.174)		(0.330)		(0.086)		(0.373)
N	71,198	71,197	71,198	71,197	71,198	71,197	71,198	71,197
Log-Lik	−1,624.930	−1,541.353	−1,965.667	−1,909.634	−1,627.715	−1,545.823	−1,830.763	−1,750.996
AIC	3,257.860	3,114.706	3,939.335	3,851.267	3265.430	3125.646	3671.525	3535.993

Note: Dependent variables are explained on page 18 and reported above each column. `Vocal Pitch` and `Overall/Congress Sentiment` are scaled to standard deviations above or below the MC's baseline. Levels of significance are reported as follows: $^{\dagger}p < .1$; $^{*}p < .05$; $^{**}p < .01$. Clustered (around each MC) standard errors are reported in parentheses.

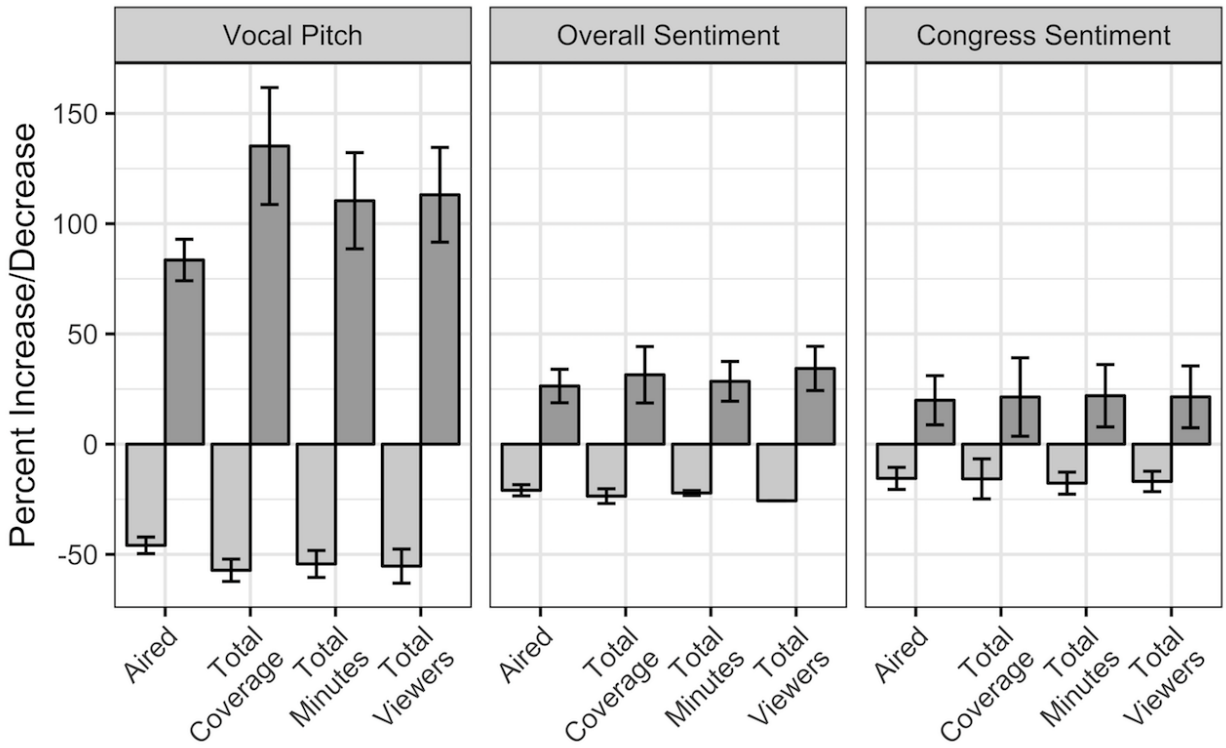
spoke one standard deviation above his/her baseline (e.g., `Vocal Pitch` = 1). We then did the same and subtracted predicted values when the speaker spoke at his/her baseline (e.g., `Vocal Pitch` = 0). Ultimately, we divided the difference using the latter, giving us the percent increase or decrease relative to the speaker's baseline. In the top-left panel, positive values imply that emotionally intense speeches – as indicated by a speaker's increased vocal pitch – are *more* likely to appear on cable news, whereas the inverse is true for negative values.

We report these percentages for two reasons. First, given our dependent variables are on different scales the raw predicted values are difficult to interpret, which is why we report a percent increase/decrease relative to the baseline. Second, we recognize and accept that most floor speeches do *not* appear on CNN, Fox News, and MSNBC. Given that, we are primarily interested in whether verbal and non-verbal emotional expressions increase the likelihood floor speeches are covered by cable news organizations. The reported percentages help answer this question, especially since they account for each variable's baseline expectation.

Figure 4 shows emotionally intense floor speeches – as indicated by a speaker's increased vocal pitch – are more likely to appear on CNN, Fox News, and MSNBC. Moving from left to right in the panel labeled "Vocal Pitch," when a MC's vocal pitch is 1 standard deviation *higher*, the speech is 83.53 percent more likely to be aired relative to what we would expect when the MC is speaking at his/her baseline. Similarly, emotionally intense speeches appear 135.24 percent more often and receive 110.42 percent more air time than we would expect given the baseline. Finally, the 113.11 percent increase in the total number of viewers suggests emotionally intense speeches are also more likely to appear on the most popular shows. These results are consistent with our first hypothesis and suggest emotionally intense floor speeches are more likely to receive television coverage.

However, emotional valence also likely matters. More specifically, news has become increasingly negative, implying speeches which express the same sentiment are also more likely to be aired. Consistent with our second hypothesis, Models 1.1 and 1.2 show floor

Figure 4: When MCs Deliver Emotionally Intense Speeches They Are More Likely to Appear on CNN, Fox News, and MSNBC



Note: Predicted values from Table 1, Models 1 (see "Aired"), 3 (see "Total Coverage"), 5 (see "Total Minutes"), and 7 (see "Total Viewers"). In the light grey bars (see), we assume the speaker spoke one standard deviation *below* his/her baseline (e.g., `Vocal Pitch` = -1). The speaker is assumed to have spoken one standard deviation *above* his/her baseline (e.g., `Vocal Pitch` = 1) in the dark grey bars (see). Both bars are standardized using predicted values when the speaker spoke at his/her baseline (e.g., `Vocal Pitch` = 0), meaning positive/negative values represent the percent increase/decrease relative to the speaker's baseline. 95 percent confidence intervals are also shown.

speeches with more negative sentiment are significantly ($p < 0.01$) more likely to be aired. Based on our results from Model 1.1, when MCs' `Overall Sentiment` is 1 standard deviation above their baseline they are 1.28 times more likely to appear on CNN, Fox News, and MSNBC as compared to floor speeches delivered at their baseline (predicted probabilities of 0.004 when `Overall Sentiment` is set to 1 and 0.003 when `Overall Sentiment` is set to 0). This holds even after accounting for party identification, ideology, institutional position, seniority, race, and whether it was an election year (see Model 1.2).

Figure 4 also shows more negative speeches – as indicated by the speaker's overall sentiment – generate *more* predicted coverage. Moving from left to right in the panel labeled "Overall Sentiment," when a MC is 1 standard deviation more *negative* in his/her speech, the speech is 26.37 percent *more* likely to be aired. Similarly, negatively valenced speeches appear 31.47 percent more often and are watched by 34.34 percent more viewers than we would expect given the baseline. Finally, the 28.49 percent increase in the total number of minutes suggests speeches are also more likely to receive air time when MCs express negativity, but this effect is 81.92 percentage points less than the comparable effect when vocal pitch is allowed to vary in a similar manner. This suggests emotional intensity – as measured by vocal pitch – has a greater influence on the quantity and quality of cable news coverage.

We next test whether this general preference for negatively valenced speeches extends to speeches where MCs speak negatively of Congress. Here, we expect to find a positive relationship, meaning when MCs express negativity towards Congress they will receive *more* cable news coverage. Indeed, MCs often run *against* Congress, meaning CNN, Fox News, and MSNBC have strong incentives to broadcast floor speeches which construct a similar narrative. Consequently, we expect to find when MCs express negativity towards Congress their speeches will receive *more* cable news coverage, especially since congressional coverage has become increasingly negative.

Consistent with our third hypothesis, Models 1.1–1.8 show when MCs express more negativity towards Congress their speeches are significantly (1) more likely to be aired, (2) more likely to appear more than once, (2) receive more air time, and (4) reach more viewers. However, the `Congress Sentiment` coefficients are only statistically significant at the 0.10-level, meaning we have less empirical support for our third hypothesis. In the SI, we also find mixed results. Although on pages S47–S47 of the SI we find `Congressional Sentiment` is consistently significant at the 0.05-level when `Vocal Pitch` and `Overall Sentiment` are not included in the same model, in other instances the variable is not as robust (e.g., see Tables S13, S14, and S16 on pages S50, S51, and S54 of the SI). Given that, we have only

partial support for our third hypothesis.

To assess the effect size, we again plotted predicted values. More specifically, in the third panel of Figure 4, we set `Congress Sentiment` to 1 in the dark grey bars and `Congress Sentiment` to -1 in the light grey bars with the former implying the speech was slightly *more* negative towards Congress than we would expect given the MC's past speeches. In both panels, the predicted values were standardized and converted to a percent increase/decrease relative to the predicted values associated with the MC's baseline. Positive values imply that increased negativity towards Congress generates *more* predicted coverage, whereas the inverse is true for negative values.

Regardless of the dependent variable, negativity towards Congress is *more* likely to appear on CNN, Fox News, and MSNBC. Moving from left to right in the panel labeled "Congress Sentiment," when a MC is 1 standard deviation more *negative* towards Congress in his/her speech, the speech is 19.91 percent *more* likely to be aired. Similarly, speeches which express negativity towards Congress are broadcast 21.39 percent more often and are watched by 21.45 percent more viewers than we would expect given the baseline. Finally, the 21.97 percent increase in the total number of minutes suggests speeches are also more likely to receive air time when MCs speak negatively about Congress.

With that said, both `Overall Sentiment` and `Congress Sentiment` have a less pronounced effect on cable news coverage as compared to `Vocal Pitch`. For example, when `Congress Sentiment` is set to 1 standard deviation above and below the speaker's baseline the predicted number of viewers decreases from 21.45 percent *above* to 16.90 percent *below* what we would expect given the MC's past speeches. When `Vocal Pitch` is allowed to vary in a similar manner, the predicted number of viewers decreases from 113.11 percent *above* to 55.32 percent *below* the speaker's baseline, suggesting the emotional intensity of a speech has a greater influence on cable news coverage as compared to other verbal measures of emotion. Indeed, when the same calculation is done for `Overall Sentiment` the predicted number of viewers goes from 34.34 percent *above* to 25.70 percent *below* the speaker's

baseline, suggesting `Vocal Pitch` is much more influential.

6 Discussion

Whether it is agenda-setting (Kingdon and Thurber, 1984), framing (Sellers, 2000), or mobilizing constituents (Wawro, 2001), television coverage helps MCs achieve their political goals (Mayhew, 1974). Unfortunately, despite the large literature on floor speeches (e.g., Maltzman and Sigelman, 1996; Harris, 2005) and congressional media coverage (e.g., Cook, 1989; Johnson and O'Grady, 2013), we know very little about the way floor speeches reach the American public. The present study addresses this noticeable gap in the literature by answering an important – yet understudied – question, what types of floor speeches are televised?

Using 74,158 floor speeches delivered in the U.S. House of Representatives and 1,974 cable news broadcasts, we not only show that floor speeches appear on the most popular programs, like *Anderson Cooper 360*, *Hannity*, *Hardball*, *Fox and Friends*, and *The Situation Room*, but some – like the speech delivered by Susan Collins on October 5, 2018 – can potentially reach millions of viewers. Indeed, we show floor speeches reach around 856,796 viewers when they are aired on cable news broadcasts, meaning a given floor speech is unlikely to appear on cable news, but those that do can reach a sizable audience. This is an important result in and of itself since most studies suggest C-SPAN is the main way floor speeches reach the American public. We show cable news is another important medium through which floor speeches can reach potential voters.

We also find speeches delivered with emotional intensity are especially likely to appear on CNN, Fox News, and MSNBC. This suggests the way a speech is delivered may be a more important predictor of television coverage as compared to what is said. Indeed, in every model we find a speaker's vocal pitch is a strong and statistically significant predictor of the quantity and quality of cable news coverage. In terms of the former, we find increased vocal pitch increases the likelihood a floor speech is aired and the number of times it is

rebroadcasted. We also find increased vocal pitch increases the total amount of air time dedicated to the speech and the speech's reach – as indicated by the total number of viewers. These results suggest audio-based measures can yield additional insights into congressional coverage and perhaps help us better understand Congress as a whole.

We also note several possible areas of future research. First, our study does not explore the nature of floor speech coverage. From an initial review of our television data, we found most speeches are used to either introduce coverage of an important debate or they are included in commentary, but several questions still remain. Does floor speech coverage vary by news outlet? Is there any variance by issue? Is the coverage of some groups, like women, fundamentally different from others? Do networks tend to show both sides of a debate to create perceived conflict? Or is there inherent partisan bias? Fortunately, the *AudioNewser* can be used to answer these and other related questions and we encourage future scholars to do so.

Second, our substantive results beg the broader question of whether MCs are intentionally "playing to the camera." Although we cannot say for sure, we suspect this is generally not the case. Not only do House press secretaries rarely employ this strategy, but it is also incredibly difficult for people to subtly manipulate their vocal inflections. More specifically, in his survey of House press secretaries Cook (1989) found "while the national networks or press may occasionally pick up a member's comment from the floor, their willingness to consider a comment catchy enough to include is impossible to predict" (98). Ultimately, this gives MCs little incentive to "play to the camera" and even in instances where "playing to the camera" may be desirable there is little evidence that MCs do not genuinely care about the issues they are advancing. For example, Sen. Collins likely knew her speech was going to be covered by CNN, Fox News, and MSNBC, but we think it is equally unlikely she strategically played up her emotions in order to garner additional coverage.

Moreover, it would be incredibly difficult – if not impossible – for Sen. Collins (or other legislators) to strategically manipulate their vocal inflections in such subtle ways. Generally

speaking, scholars think of verbal and nonverbal behavior as part of a "leakage hierarchy" with "verbal content" (i.e., the words spoken) located in the "controllable end of the continuum, whereas the body and tone of voice may be classified as less controllable and more leaky channels" (Zuckerman and Driver, 1985, 130). This not only suggests it is more difficult to control one's voice, but individuals who attempt to control their vocal pitch, often sound "more tense and less pleasant or compelling than someone speaking sincerely," which is in turn associated with "increased vocal pitch" (Elkins et al., 2014, 505).

Sen. Collins likely did not come up with "gutter-level political campaign" off the cuff, but to say that she strategically increased her vocal pitch by 12 Hz to increase the probability her speech appeared on CNN, Fox News, and MSNBC by some unknown amount seems unlikely. We think a much more likely explanation is that MCs genuinely care about some issues more than others, meaning when they speak about those issues they tend to speak with greater emotional intensity. Journalists are likely drawn to those speeches, not because they prefer speeches delivered 12Hz higher, but because they prefer speeches in which speakers sound like they are emotionally invested in the points they are advancing. The present paper uses vocal pitch as one way to measure this concept, but this is by no means the only way and we encourage future scholars to explore other ways to quantify the emotional intensity of speech.

With that said, the Democratic Message Board and the Republican Theme Team were created to coordinate collective speaking efforts, meaning the House floor is undoubtedly used strategically (Harris, 2005). Rep. Newt Gingrich's (R-GA) use of one-minute speeches in order to attack Democrats and advance the Republican agenda is also a well-known example of the strategic use of floor speeches (Maltzman and Sigelman, 1996). Even though we think it is unlikely that MCs are intentionally manipulating their vocal pitch, we fully acknowledge political parties strategically use floor speeches in this way. These efforts not only underline the perceived importance of floor speeches being picked up by news organizations, but they also offer another important area of future research.

According to Cook (1989), "the value of broadcasting floor proceedings is not C-SPAN's audiences around the country but the networks' feeds or the reporters who tune in" (99). For example, when Republicans took to the floor in early 1984 to address issues Democrats were bottling up in committees, "Gingrich estimated 200,000 viewers were tuned into C-SPAN at any one time" (Cook, 1989, 99). However, after the Republican speaking effort lead to an outburst by Speaker Tip O'Neill (D-MA), the "ensuing publicity" suggested Gingrich was less interested in C-SPAN's audience than "getting through to the press" (Cook, 1989, 100). Understanding such collective speaking efforts is especially important given the polarizing effect of cable news broadcasts (Levendusky, 2013*b*) and the two-step flow of communication (Levendusky, 2013*a*). Fortunately, future scholars can use the *AudioNewser* to answer these and other questions related to what appears on CNN, Fox News, and MSNBC.

C-SPAN has consistently been offered as the primary way by which floor speeches reach voters, but we show floor speeches increasingly appear on cable news broadcasts. Even though we find the words influence whether a speech is aired on CNN, Fox News, or MSNBC, the way those words are delivered seems to carry more weight. While text has long been a focal point of political science research, we demonstrate that non-verbal expressions may yield additional insights. In doing so, we provide an important foundation for future research in which text, audio, and video data are used together in order to better understand the causes and consequences of legislative speech and behavior more broadly. Our software – the *AudioNewser* – makes such research more accessible while simultaneously expanding the way we study cable news.

References

Bennett, W Lance, and Robert M Entman. 2001. "Mediated politics: An introduction." *Mediated politics: Communication in the future of democracy* 1: 1–30.

Berger, Jonah, and Katherine L Milkman. 2012. "What makes online content viral?" *Journal*

of marketing research 49(2): 192–205.

Bradley, Margaret M, Mark K Greenwald, Margaret C Petry, and Peter J Lang. 1992. "Remembering Pictures: Pleasure and Arousal in Memory." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18(2): 379–390.

Burgoon, Judee K. 1978. "Attributes of the Newscaster's Voice as Predictors of His Credibility." *Journalism Quarterly* 55(2): 276–300.

Caple, Helen, and Monika Bednarek. 2013. "Delving into the discourse: Approaches to news values in journalism studies and beyond.".

Cappella, Joseph N, and Kathleen Hall Jamieson. 1996. "News frames, political cynicism, and media cynicism." *The Annals of the American Academy of Political and Social Science* 546(1): 71–84.

Cohen, Jeffrey E. 2008. *The presidency in the era of 24-hour news.* Princeton University Press.

Cook, Timothy. 1989. *Making Laws and Making News: Media Strategies in the U.S. House of Representatives.* Washington, D.C.: Brookings Institution Press.

Cotton, Courtenay V, and Daniel PW Ellis. 2010. "Audio fingerprinting to identify multiple videos of an event." Paper presented at the Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on,.

Dietrich, Bryce J., Ryan D. Enos, and Maya Sen. 2018. "Emotional Arousal Predicts Voting on the U.S. Supreme Court." *Political Analysis* , to appear.

Ekman, Paul, Maureen O'Sullivan, Wallace V. Friesen, and Klaus R. Scherer. 1991. "Invited Article: Face, Voice, and Body in Detecting Deceit." *Journal of Nonverbal Behavior* 15(2): 125–135.

Elkins, Aaron, Stefanos Zafeiriou, Maja Pantic, and Judee Burgoon. 2014. "Unobtrusive Deception Detection." In *The Oxford Handbook of Affective Computing*, eds. Rafael A. Calvo, Sidney D'Mello, Johnathan Gratch, and Arvid Kappas. New York, NY: Oxford University Press , 503–515.

Fenno, Richard F. 1977. "U.S. House Members in Their Constituencies: An Exploration." *American Political Science Review* 71(3): 883–917.

Forgette, Richard, and Jonathan S Morris. 2006. "High-conflict television news and public opinion." *Political Research Quarterly* 59(3): 447–456.

Galtung, Johan, and Mari Holmboe Ruge. 1965. "The structure of foreign news: The presentation of the Congo, Cuba and Cyprus crises in four Norwegian newspapers." *Journal of peace research* 2(1): 64–90.

Geer, John G. 2008. *In defense of negativity: Attack ads in presidential campaigns.* University of Chicago Press.

Gershon, Sarah. 2012. "When race, gender, and the media intersect: Campaign news coverage of minority congresswomen." *Journal of Women, Politics & Policy* 33(2): 105–125.

Grabe, Maria Elizabeth, Shuhua Zhou, and Brooke Barnett. 2001. "Explicating sensationalism in television news: Content and the bells and whistles of form." *Journal of Broadcasting & Electronic Media* 45(4): 635–655.

Grabe, Maria Elizabeth, Shuhua Zhou, Annie Lang, and Paul David Bolls. 2000. "Packaging Television News: The Effects of Tabloid on Information Processing and Evaluative Responses." *Journal of Broadcasting & Electronic Media* 44(4): 581–598.

Graber, Doris A. 1994. "The infotainment quotient in routine television news: A director's perspective." *Discourse & Society* 5(4): 483–508.

Graber, Doris A. 1996. "Say it with pictures." *The annals of the American academy of political and social science* 546(1): 85–96.

Graber, Doris A, and Johanna Dunaway. 2017. *Mass Media and American Politics.* CQ Press.

Hamilton, James. 2004. *All the News That's Fit to Sell: How the Market Transforms Information into News.* Chicago: Princeton University Press.

Harcup, Tony, and Deirdre O'Neill. 2016. "What is news?" *Journalism Studies* .

Harris, Douglas B. 2005. "Orchestrating Party Talk: A Party-Based View of One-Minute Speeches in the House of Representatives." *Legislative Studies Quarterly* 30(1): 127–141.

Imai, Kosuke, Gary King, and Olivia Lau. 2008. "Toward A Common Framework for Statistical Analysis and Development." *Journal of Computational Graphics and Statistics* 17(4): 892–913.

Johnson, Tyler, and Caitlin O'Grady. 2013. "Speakers and the Spotlight: Explaining Media Coverage of Leadership in the House of Representatives." *American Politics Research* 41(3): 498–532.

Kim, Hyun Suk. 2015. "Attracting views and going viral: How message features and news-sharing channels affect health news diffusion." *Journal of Communication* 65(3): 512–534.

King, Gary, and Langche Zeng. 2001. "Logistic Regression in Rare Events Data." *Political Analysis* 9(2): 137–163.

Kingdon, John W, and James A Thurber. 1984. *Agendas, alternatives, and public policies.* Vol. 45 Little, Brown Boston.

Kuklinski, James H, and Lee Sigelman. 1992. "When Objectivity is Not Objective: Network Television News Coverage of US Senators and the "Paradox of Objectivity"." *The Journal of Politics* 54(3): 810–833.

Lang, Annie, Kulijinder Dhillon, and Qingwen Dong. 1995. "The Effects of Emotional Arousal and Valence on Television Viewers' Cognitive Capacity and Memory." *Journal of Broadcasting & Electronic Media* 39(3): 313–327.

Levendusky, Matthew. 2013*a*. *How partisan media polarize America*. University of Chicago Press.

Levendusky, Matthew S. 2013*b*. "Why do partisan media polarize viewers?" *American Journal of Political Science* 57(3): 611–623.

Lichter, R Stephen, and Daniel R Amundson. 1994. "Less news is worse news: Television coverage of Congress, 1972-92." Paper presented at the Congress, the Press, and the Public,.

Maltzman, Forrest, and Lee Sigelman. 1996. "The Politics of Talk: Unconstrained Floor Time in the U.S. House of Representatives." *The Journal of Politics* 58(3): 819–830.

Martín-Santana, Josefa D, Eva Reinares-Lara, and Pedro Reinares-Lara. 2017. "How Does the Radio Spokesperson's Voice Influence Credibility?" *Communications* 42(2): 151–172.

Mauss, Iris B., and Michael D. Robinson. 2009. "Measures of Emotion: A Review." *Cognition and Emotion* 23(2): 209–237.

Mayhew, David R. 1974. *Congress: The Electoral Connection*. New Haven, CT: Yale University Press.

Pantti, Mervi. 2010. "The value of emotion: An examination of television journalists? notions on emotionality." *European Journal of Communication* 25(2): 168–181.

Patterson, Thomas E. 2011. *Out of Order: An incisive and boldly original critique of the news media's domination of American Political Process.* Vintage.

Peters, Chris. 2011. "Emotion aside or emotional side? Crafting an ?experience of involvement?in the news." *Journalism* 12(3): 297–316.

Reese, Stephen D, Tim P Vos, and Pamela J Shoemaker. 2009. "Journalists as gatekeepers." In *The handbook of journalism studies.* Routledge , 93–107.

Roberts, Margaret E., Brandon M. Stewart, and Dustin Tingley. 2014. "stm: R package for structural topic models." *R Package* 1: 1–12.

Robinson, Michael J, and Kevin R Appel. 1979. "Network News Coverage of Congress." *Political Science Quarterly* 94(3): 407–418.

Rozell, Mark J. 1994. "Press coverage of Congress 1946–1992." Paper presented at the Congress, the Press, and the Public,.

Sellers, Patrick J. 2000. "Manipulating the Message in the US Congress." *Harvard International Journal of Press/Politics* 5(1): 22–31.

Shoemaker, Pamela J, and Tim P Vos. 2009. *Gatekeeping theory.* Routledge.

Soroka, Stuart N. 2012. "The gatekeeping function: Distributions of information in media and the real world." *The Journal of Politics* 74(2): 514–528.

Sparrow, Bartholomew H. 1999. *Uncertain Guardians: The News Media as a Political Institution.* Baltimore: Johns Hopkins University Press.

Uribe, Rodrigo, and Barrie Gunter. 2007. "Are Sensational News Stories More Likely to Trigger Viewers' Emotions than Non-Sensational News Stories? A Content Analysis of British TV News." *European Journal of Communication* 22(2): 207–228.

Vogel, Adam P., Paul Maruff, Peter J. Snyder, and James C. Mundt. 2009. "Standardization of Pitch Range Settings in Voice Acoustic Analysis." *Behavior Research Methods* 41(2): 318–324.

Vos, Debby. 2014. "Which Politicians Pass the News Gates and Why? Explaining Inconsistencies in Research on News Coverage of Individual Politicians." *International Journal of Communication* 8: 2438–2461.

Wahl-Jorgensen, Karin. 2013. "The strategic ritual of emotionality: A case study of Pulitzer Prize-winning articles." *Journalism* 14(1): 129–145.

Waismel-Manor, Israel, and Yariv Tsfati. 2011. "Why do better-looking members of congress receive more television coverage?" *Political Communication* 28(4): 440–463.

Warhurst, Samantha, Patricia McCabe, and Catherine Madill. 2013. "What Makes a Good Voice for Radio: Perceptions of Radio Employers and Educators." *Journal of Voice* 27(2): 217–224.

Wawro, Gregory. 2001. *Legislative entrepreneurship in the US House of Representatives*. University of Michigan Press.

Zuckerman, Miron, and Robert E. Driver. 1985. "Telling Lies: Verbal and Nonverbal Correlates of Deception." In *Multichannel Integrations of Nonverbal Behavior*, eds. Aaron W. Siegman, and Stanley Feldstein. New York, NY: Lawrence Erlbaum , 129–147.

Supporting Information (to go online) for: This Floor Speech Will Be Televised: Understanding the Factors that Influence When Floor Speeches Appear on Cable Television.

Contents

S1 Description of Floor Speech Data

In this section we describe the floor speech data we collected from *HouseLive*. We also discuss our approach to extracting the text and vocal pitch from each speech.

S1.1 House Video Archives

We used *HouseLive*[S1] primarily because they make their audio and video data publicly available. The website provides both `mp3` and `mp4` files for each legislative session. We used the former to expedite the download time. To download these files, please use the `R` script found here:

```
URL removed for blind review.
```

For those interested in downloading the `mp4` files, please use this `R` script:

```
URL removed for blind review.
```

We downloaded 863 `mp3` files spanning from January 6, 2009 to August 4, 2014 which took around 10 hours.

S1.2 Extracting Floor Speeches

Closed-captioning information was used to identify the speakers and floor speech text. Unlike the *Congressional Record*, closed-captioning information has the advantage of reporting verbatim what is said on the House floor. For example, MCs often read things into the *Congressional Record*, in which case the document will be added to the *Congressional Record* without ever being spoken on the floor. Since close-captions are transcribed in real-time they do not suffer from the same limitation which is why we used them for this study.

When using closed-captioning information, typographical errors are a legitimate concern. To address this issue we contacted the company that provides the closed-captioning service for the House of Representative. They told us that their internal validation checks yield a 5 percent error rate, meaning 5 percent of the time the closed-captioning does not align with what is being said, whereas the inverse is true 95 percent of the time. To check, we transcribed 100 randomly selected speeches and compared the manual transcripts with the close-captions using a variety of similarity metrics. Ultimately, we found little difference which made us confident the closed-captioning found on *HouseLive* is an accurate reflection of what is said in the U.S. House of Representatives.

The individual `mp3` files we collected averaged 7 hours and 27 minutes in length. We used `ffmpeg` to split these longer audio files into individual speeches using the closed-captioning

[S1] `http://houselive.gov/`

time stamps. This resulted in 152,117 `wav` files. Since many of these speeches were incredibly short and contained very little policy content, we restricted our analysis to those speeches that had at least 50 words. This resulted in text and audio for 74,158 speeches. We provide a short working example here:

```
URL removed for blind review.
```

The colon after a speaker's name (e.g., PELOSI:) was used to identify which MC was speaking. Using `ffmpeg`, we then used the associated timestamps to split the longer `mp3` files into individual files for each speech. Again, we provide a short working example here:

```
URL removed for blind review.
```

S1.3 Extracting Vocal Pitch

Vocal pitch was extracted using *Praat*[S2] which implements an algorithm developed by Boersma (1993). Generally speaking, this popular software estimates the fundamental frequency by dividing the autocorrelation of a windowed signal by the autocorrelation of the window itself. To use the software, one has to set five parameters: the pitch floor, pitch ceiling, window length, window shape, and voicing threshold. As suggested by *Praat*, for male MCs we set the pitch floor and ceiling to 75Hz and 300Hz, respectively. For female MCs, the pitch range was set to 100-500Hz. Examples of the resulting output can be found here:

```
URL removed for blind review.
```

In all models, we standardized vocal pitch. This was done for three reasons. First, because women's vocal cords tend to be smaller and shorter they typically speak at a higher vocal pitch than men. Once vocal pitch is scaled to standard deviations above or below each speaker's baseline (or mean) vocal pitch, we account for these inherent sex differences while "canceling out" potential irregularities in the pitch contour (Hess 2007). Moreover, we are not really interested in whether news organizations are more or less likely to broadcast MCs who generally speak at a higher vocal pitch. Rather, we want to know whether relative increases in emotional intensity – as indicated by a speaker's vocal pitch – generate more television coverage. Our standardization approach is well-suited to assess this relationship.

S2 Description of Television Data

In this section we describe the *Internet Archive* television data. Please refer to Section 4 in main text for more details. Figure 2 in the main text graphically displays the workflow we describe in this section of the Supplemental Information.

[S2] `http://www.fon.hum.uva.nl/praat`

S2.1 Recording Television

The television data is drawn from the *TV News Archive*.[S3] The original data was created by splitting several cable feeds into minute-long segments for each program using a known television schedule. These are then stored on local servers. Once compiled into programs ranging from 30 minutes to 12 hours, the *Archive* then converts the files into a number of formats, including `mp4` video files.

Unfortunately, due to copyright issues we cannot provide the actual `mp4` files we used for this study, but we have provided a few examples:

```
URL removed for blind review.
```

We have also included some sample scripts which show how the one-minute segments were combined into the larger `mp4` files. The example code can be found here:

```
URL removed for blind review.
```

S2.2 Audio Fingerprinting

An audio fingerprint is a summarized version of an audio file, one that has removed everything except the most distinctive features of every few milliseconds. Undoubtedly, there are a number of ways to summarize a piece of audio. For this study, we use an algorithm originally developed by Dan Ellis and implemented in *audfprint*.[S4]

As explained on pages 10–12 in the main text, the algorithm involves the following steps:

1. Estimate the frequencies associated with the most prominent waves in a given audio file
2. Using single-second intervals, identify the local maxima in the associated spectrograms
3. Compare the local maxima with the nearest neighbors
4. A "landmark" is recorded when the pairs of maxima are sufficiently distinguishable from the local maxima.
5. A "sufficient" difference is determined using binned maxima from the second and third peaks, as well as the number of time steps between all three peaks.
6. A "hash table" is created using the output for a given audio file. The landmarks are converted into 32-bit strings and stored in buckets associated with unique hashes.

[S3] `https://archive.org/details/tv`
[S4] `https://github.com/dpwe/audfprint`

7. These buckets can then be searched using the landmarks from another audio file.
8. The "search" involves comparing the hash obtained from the new audio file with every hash stored in the hash table.

This process was repeated for all the audio files in our floor speech and television data. The hash tables for both the floor speech and television data are included in the original software – called *AudioNewser* – we created for this study.

S2.3 *AudioNewser*

The *AudioNewser* allows users to query both the floor speech and television data using their own audio files. Our software can be downloaded here:

`URL removed for blind review.`

Please refer to the `README` file for installation instructions. To run our software the following programs need to be installed:

- audfprint
- ffmpeg
- Python 3

Ultimately, our software is very similar to *Shazam*. Readers can use the main function to query our television data using their own audio files, such as those obtained from advertisements or videos posted on social media. A companion *YouTube* tutorial can be found here:

`URL removed for blind review.`

We have also posted a working example on *YouTube*. That video can be found here:

`URL removed for blind review.`

The code is written to run on Unix-based operating systems. Please contact the corresponding author for Windows support.

S2.4 *Storyboard Tool*

In order to help users verify the *AudioNewser* we created an additional tool which returns an automatically generated storyboard for each news broadcast. Although there are a variety of ways to achieve this end (e.g., Sheena and Narayanan 2015; Zhuang et al. 1998), the alpha version of this tool uses closed-captioning breaks to divide the longer video into a number of "scenes." Future iterations of the software will likely use a variation of the algorithm developed by Kim and Hwang (2002). The alpha software can be found here:

`URL removed for blind review.`

Please consult the `README` file for installation instructions. Again, a companion *YouTube* tutorial can be found here:

`URL removed for blind review.`

We have also posted a working example. That video can be found here:

`URL removed for blind review.`

The code is written to run on Unix-based operating systems. Please contact the corresponding author for Windows support.

S3 Description of Sentiment Data

In this section we describe the sentiment data. Please refer to pages 14–17 in the main text for more details. In this portion of the Supplemental Information, we describe the basic workflow.

S3.1 `analyzeSentiment`

Similar to other REST APIs, the Google Cloud Natural Language API (Google API)[S5] allows users to extract information about people, places, and things mentioned in text documents, ranging from newspaper articles to blog posts. The `analyzeSentiment` function identifies the prevailing emotional opinion within the text, especially whether the writer's attitude was positive, negative, or neutral.

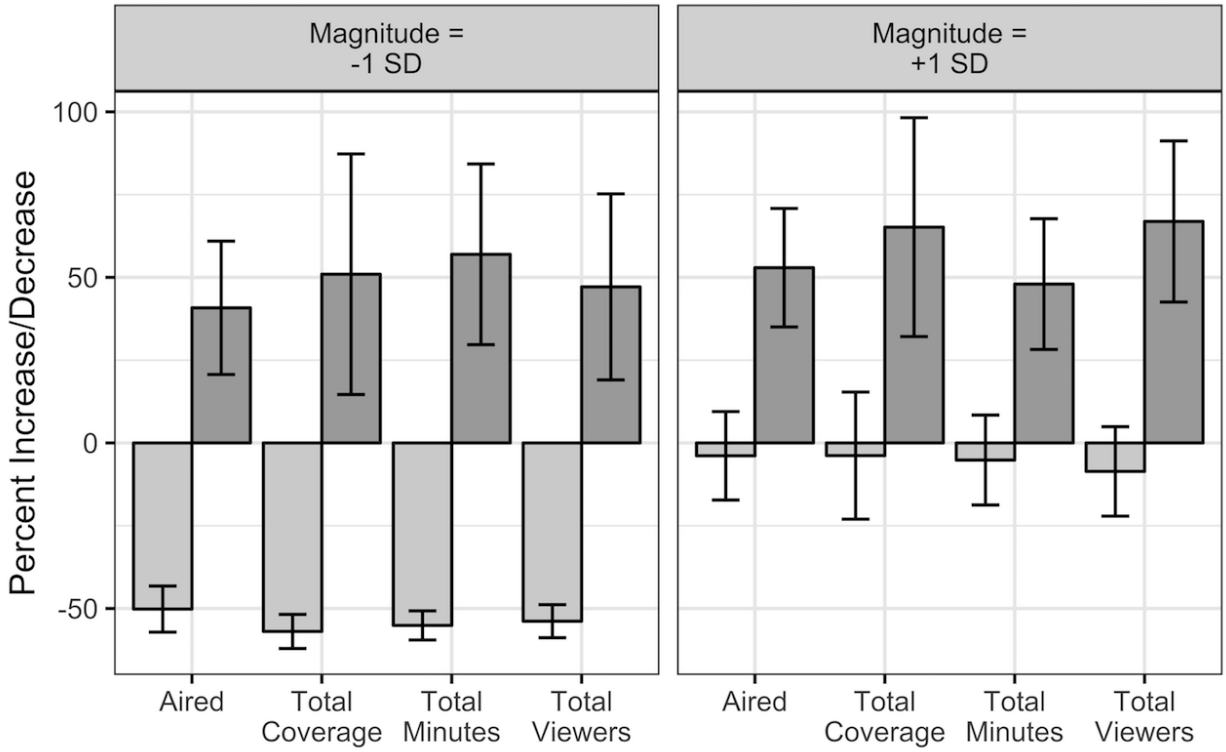
The Google API ultimately returns a sentiment score and magnitude value. A document's *score* indicates the overall emotion of a document. The *magnitude* indicates how much

[S5] `https://cloud.google.com/natural-language/`

emotional content is present within the document which is often proportional to the length of the document. The Google API only determines whether the emotion is generally positive or negative. It does not identify specific positive and negative emotions. For example, "angry" and "sad" are both considered negative emotions, meaning the Google API will essentially return the same result when a MC expresses either emotion.

For the purposes of this study, we did not use the magnitude scores. In Table S1 we report results from base models where the speech's overall sentiment and magnitude are interacted. Both variables are standardized using an MC's past speeches. Regardless of the dependent variable, the main effects of sentiment and magnitude are statistically significant, but the interaction term is mostly indistinguishable from zero.

Figure S1: Predicted Values for Table S1



Note: Predicted values from Table S1, Models 1 (see "Aired"), 2 (see "Total Coverage"), 3 (see "Total Minutes"), and 4 (see "Total Viewers"). `Overall Magnitude` set to -1 and 1 in the left and right panels, respectively. In the light grey bars (see ▢), `Overall Sentiment` was set to -1 in the light grey bars, meaning the MC expressed *less* negativity. `Overall Sentiment` was set to 1 in the dark grey bars (see ■), meaning the MC expressed *more* negativity. The light and dark grey bars were standardized using predicted values when `Overall Magnitude` and `Overall Sentiment` were both set to 0, meaning positive/negative values represent the percent increase/decrease relative to the speaker's baseline. Error bars represent 95 confidence intervals.

Table S1: When MCs Deliver More Negative Speeches They Are More Likely to Appear on CNN, Fox News, and MSNBC (with Magnitude)

	Dependent variable:			
	Aired	Total Coverage	Total Minutes	Total Viewers
	(1)	(2)	(3)	(4)
Constant	-5.730^{**}	-5.097^{**}	-3.510^{**}	-16.208^{**}
	(0.066)	(0.142)	(0.441)	(1.836)
Overall Sentiment	0.387^{**}	0.454^{**}	0.168^{**}	0.790^{**}
	(0.055)	(0.086)	(0.036)	(0.150)
Overall Magnitude	-0.191^{*}	-0.226^{*}	$-0.065^{\dagger}$	-0.356^{*}
	(0.078)	(0.095)	(0.036)	(0.175)
Overall Sentiment × Overall Magnitude	0.151^{*}	$0.184^{\dagger}$	$0.077^{\dagger}$	0.237
	(0.075)	(0.105)	(0.044)	(0.156)
N	74,146	74,146	74,146	74,146
Log Lik	−1,699.163	−2,034.802	−1,703.653	−1,901.521
AIC	3,406.327	4,077.603	3,417.307	3,813.042

Note: The dependent variable is shown above the corresponding columns. These are (1) whether a floor speech appeared on television, (2) the number of times the floor speech was televised, (3) the total air time (in minutes), and (4) the estimated number of total viewers (in millions). Model 1 was estimated using a rare event logistic regression. Models 2 was estimated using a negative binomial regression. Models 3 and 4 were estimated using a Tobit regression left-censored at 0. `Overall Sentiment` and `Overall Magnitude` are scaled to standard deviations above or below an MC's baseline. In terms of the former, positive values imply the MC expressed more negativity, whereas in the latter positive values imply the MC dedicated a greater proportion of his/her speech to emotional content. Levels of significance are reported as follows: $^{\dagger}p < .1$; $^{*}p < .05$; $^{**}p < .01$. Clustered (around each MC) standard errors are reported in parentheses.

Figure S1 reports predicted values from Table S1, Models 1 (see "Aired"), 3 (see "Total Coverage"), 5 (see "Total Minutes"), and 7 (see "Total Viewers"). We set overall sentiment to 1 in the dark grey bars, implying the speech was slightly *more* negative than we would expect given the MC's past speeches. In the light grey bars, we set the overall sentiment to -1, implying the speech was slightly *less* negative than we would expect given the MC's past speeches. Magnitude is set to 1 in the panel furthest to the right. This implies the speaker dedicated a greater percentage of his/her speech to emotional content than we would expect given his/her baseline (see "Magnitude = +1 SD"). The inverse is true for the panel furthest to the left. Here, magnitude was set to -1, meaning the speaker dedicated less of his/her speech to emotional content than we would expect given his/her past speeches. Ultimately, all predicted values were standardized to a percent increase/decrease relative to the predicted values obtained when sentiment and magnitude were both set to 0.

We found the difference between the dark and light grey bars *increases* when speeches contain more emotional content. More specifically, when magnitude is set to 1, the average difference between predicted values when the floor speech express a negative (" Overall Sentiment" = 1) as opposed to positive ("Overall Sentiment" = -1) sentiment is 103.73 percent, suggesting a strong preference for negatively valanced speeches. For example, floor speeches are 40.92 percent *more* likely to be aired (as compared to the baseline) when a negative sentiment is expressed at a *higher* magnitude (see "Aired" in right panel). The likelihood decreases 92.00 percentage points to -51.07 percent when overall sentiment is set to -1, suggesting when a speech contains a lot of emotional content it is *less* likely to be aired when a positive sentiment is expressed (see "Aired" in right panel).

The same result is found in the left panel, but the difference between the dark and light grey bars is smaller when less of the speech is dedicated to emotional content. More specifically, when magnitude is set to -1, the average difference between predicted values when the floor speech expresses a negative ("Overall Sentiment" = 1) as opposed to positive ("Overall Sentiment" = -1) sentiment is 64.01, suggesting a weaker preference for negatively valanced speeches. For example, floor speeches are 52.75 percent more likely to be aired (as compared to the baseline) when a negative sentiment is expressed towards Congress at a *lower* magnitude (see "Aired" in left panel). The likelihood decreases another 55.97 percentage points to 3.22 percent below the speaker's baseline when overall sentiment is set to -1, suggesting when a speech contains less emotional content it is *less* likely to be aired when a positive sentiment is expressed towards Congress (see "Aired" in left panel).

Ultimately, the average difference between the dark and light grey bars is 1.62 times higher when a speech is delivered with more emotional content. For example, when a floor speech expresses a negative ("Overall Sentiment" = 1) as opposed to positive ("Overall Sentiment" = -1) sentiment the likelihood it is aired decreases 103.73 percentage points when the speech contains a lot of emotional content ("Magnitude" = 1). When the speech contains less emotional content ("Magnitude" = -1), the likelihood only decreases 64.01 percentage points, suggesting the preference for negatively valanced speeches is *more* pronounced when a higher proportion of the speech is dedicated to emotional content. Indeed, when the av-

erage difference between the light and dark grey bars in the left panel (64.01) is compared to the average difference in the right panel (103.73) using a simple t-test, one finds a strong and statistically significant difference between the two means ($t = 5.71, df = 6, p < 0.01$).

Given that the results are very similar to those reported in the main text, we did not include the magnitude scores in our main analysis. This was ultimately done to make Table 1 more interpretable. It is difficult enough to interpret three independent variables, especially given the limited number of floor speeches that actually appear on CNN, Fox News, and MSNBC. We concluded the added benefit was outweighed by the decreased degrees of freedom. Table S1 and Figure S1 suggest the results look largely the same when magnitude is included, so we think this was a reasonable choice.

S3.2 `analyzeEntitySentiment`

The Google API also provides information about entities in the text, which generally refer to named "things" such as famous individuals, landmarks, common objects, etc. According to the Google API documentation, entities broadly fall into two categories: proper nouns that map to unique entities (specific people, places, etc.) or common nouns (also called "nominals" in natural language processing). Generally speaking, if something is a noun, it qualifies as an "entity."

The `analyzeEntitySentiment` function determines the sentiment (positive or negative) expressed about entities within the text. Similar to the overall sentiment, a score and magnitude value is returned for each entity reference. Those scores and values are then aggregated into the overall sentiment score and magnitude for an entity. We report the ten entities that appeared the most in our data in Table S2. The last two columns report the associated sentiment and magnitude. Ultimately, we found the latter did not significantly alter the results, so we only used the overall sentiment scores for the present paper.

The metadata used for the database of searchable entities is compiled from a variety of sources, including Wikipedia articles and Google searches. Although the number of keywords undoubtedly influences the likelihood an entity appears, the Google API is not returning simple regular expression counts. This is shown in the third column of Table S2. As you can see, the number of times an entity appears does not correspond with the number of times the entity label appears within our data. Given that, it is clear the Google API is more than a keyword search. Instead, entities are extracted and sentiment is analyzed using natural language processing and Google's algorithm. For more information on named entity recognition, please refer to Nadeau and Sekine (2007).

S3.3 Extracting Positive/Negative Sentiment

The `Python` code we used to apply the `analyzeSentiment` and `analyzeEntitySentiment` functions to our floor speech data can be found in the zip file found here:

Table S2: Top-10 Entities

Label	Entity Count	Keyword Count	Sentiment	Magnitude
`people`	99,355	98,572	-0.05	0.25
`bill`	78,450	98,636	-0.02	0.24
`speaker`	55,050	67,336	-0.01	0.15
`country`	44,915	44,784	0.00	0.22
`congress`	41,364	42,427	-0.04	0.20
`americans`	36,216	28,193	-0.05	0.28
`government`	35,416	46,046	-0.05	0.23
`jobs`	34,610	40,356	-0.03	0.24
`amendment`	32,844	35,065	-0.04	0.30
`states`	32,199	41,635	-0.02	0.15

Note: The ten entities that appeared the most in our floor speech data can be found in the first column. The number of appearances is reported in the second column. The third column reports the number of times the entity label appears in the data using a simple keyword search. The entity sentiment and magnitude are reported in the fourth and fifth columns, respectively.

```
URL removed for blind review.
```

Please consult the `README` file for detailed instructions. The entity results for our floor speech data can be found here:

```
URL removed for blind review.
```

To query the entity results, please use the following code:

```
URL removed for blind review.
```

The code is written to run on Unix-based operating systems. Please contact the corresponding author for Windows support.

S3.4 Comparing `analyzeEntitySentiment` to Dictionary-Based Measures

S3.4.1 "Positive" and "Negative" Words

We used seven dictionaries to validate the sentiment measures we obtained from the Google API. The Linguistic Inquiry Word Count (LIWC) dictionary can be purchased online – `http://liwc.wpengine.com` – for $89. For this study, we used the 2007 version of the

dictionary. In total, LIWC includes 407 "positive" emotion words, including their extensions. Here are some examples of the positive words included in the dictionary:

happy

pretty

good

In addition to these "positive" words, LIWC includes 500 "negative" emotion words, including their extensions. Here are some examples of the negative words included in the dictionary:

hate

worthless

enemy

The Harvard IV-4 dictionary is publicly available (`http://www.wjh.harvard.edu/~inquirer/homecat.htm`) and contains 1,915 and 2,291 "positive" and "negative" words, respectively. Here are some examples of the positive words included in the dictionary:

adorable

compassionate

enjoyable

Some examples of the negative words found in the Harvard IV-4 dictionary can be found here:

abrasive

condescending

derisive

We could not find the Dictionary of Affective Language (DAL) online. We also could not obtain the dictionary from the author Cynthia Whissell. The best we could do is obtain a list of all the words included in the dictionary, regardless of category (pleasantness, activation, or imagery). Once we had the list, we looked for words that appeared in both the DAL and LIWC. Words that appeared in DAL and the LIWC positive emotion words category were said to be "positive." Words that appeared in DAL and the LIWC negative emotion words category were said to be "negative." Here are some examples of the positive words included in the our version of DAL:

cheerful

delighted

enthusiastic

Some examples of the negative words found in our version of DAL can be found here:

alarming

disgusting

enrage

Unfortunately, for copyright reasons we cannot provide the DAL or LIWC word lists. However, to give some comparison, we provide all the "positive" and "negative" words we used from the Harvard IV-4 dictionary:

`URL removed for blind review.`

The `afinn` dictionary was developed by Nielsen (2011) and can be found in the `tidytext` package in the `R` Statistical Language. It can also be installed using the `Python` module posted on the author's *GitHub* page (`https://github.com/fnielsen/afinn`). Here are some examples of the positive words (or words with sentiment scores greater than zero) found in the `afinn` dictionary:

amuse

positively

lively

Here are some examples of the negative words (or words with sentiment scores less than zero) found in the `afinn` dictionary:

denounce

threats

sorrowful

The `bing` dictionary was developed by Prof. Bing Liu from the University of Illinois-Chicago and can also be found on his website (`http://www.cs.uic.edu/~liub/FBS/opinion-lexicon-English` `rar`). Here are some examples of the positive words found in the `bing` dictionary:

faithfully

gleeful

tremendously

We also prove some examples of the negative words found in the `bing` dictionary:

concerned

frighten

toxic

The `loughran` dictionary was developed by Prof. Tim Loughran and Bill McDonald from the University of Notre Dame and can also be found online (`https://sraf.nd.edu/textual-analysis/resources/#LM%20Sentiment%20Word%20Lists`). Here are some examples of the positive words found in the `loughran` dictionary:

favorable

pleasure

successfully

We also prove some examples of the negative words found in the `loughran` dictionary:

criticizes

defaced

wasted

Finally, the `nrc` dictionary was developed by Saif M. Mohammad from the National Research Council (Canada) and can also be found on his website (`https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm`). Here are some examples of the positive words found in the `nrc` dictionary:

entertain

prosper

sentimental

We also prove some examples of the negative words found in the `nrc` dictionary:

hopelessness

jealousy

unimpressed

Interested readers can get the full word lists using the `tidytext` package. For example, to get all of the positive words from the `nrc` dictionary use this `R` code:

```
get_sentiments("nrc")$word[get_sentiments("nrc")$sentiment=="positive"]
```

S3.4.2 Dictionary-Based Measures Sentiment

To create the overall sentiment measures for Figure 3 in the main text, we simply applied the aforementioned dictionaries to each floor speech. We then took the percentage of positive words and subtracted the percentage of negative words. Although not the same as the Google API's measure, we think this is the closest approximation given the dictionaries we acquired. To validate our measure, we then calculated the correlation between each dictionary-based measure and the results from the Google API (see dark gray bars in Figure 3).

The dictionary-based congressional sentiment measures were created using a similar process, but instead of applying the dictionaries to the whole speech we applied the dictionaries to the 10 words surrounding the following words and phrases: "Congress," "United States House of Representatives," "Senate," "United States Senate," and "Capitol Hill." For example, consider this portion of Sen. Collins' speech on Justice Kavanaugh:

> Another assertion that I have heard often that Judge Kavanaugh cannot be trusted if a case involving alleged wrongdoing by the president were to come before the court. The basis for this argument seems to be two-fold. First, Judge Kavanaugh has written that he believes that Congress should enact legislation to protect presidents from criminal prosecution or civil liability while in office. Mr. President, I believe opponents missed the mark on this issue. The fact that judge Kavanaugh offered this legislative proposal suggests that he believes that the president does not have such protection currently.

After we applied some common pre-processing steps (converted to lower case, removed punctuation, removed numbers, and removed stop words), this section of her speech becomes:

> another assertion heard often judge kavanaugh trusted case involving alleged wrongdoing president come court basis argument seems twofold first judge kavanaugh written believes congress enact legislation protect presidents criminal

> prosecution civil liability office mr president believe opponents missed mark issue fact judge kavanaugh offered legislative proposal suggests believes president protection currently

We then identified the key words and phrases (highlighted in blue) and the ten surrounding words (highlighted in red) using a regular expression:

> another assertion heard often judge kavanaugh trusted case involving alleged wrongdoing president court basis argument seems twofold first judge kavanaugh written believes congress enact legislation protect presidents criminal prosecution civil liability office mr president believe opponents missed mark issue fact judge kavanaugh offered legislative proposal suggests believes president protection currently

We would then applied each dictionary and calculated the percent of positive and negative words using the reduced word list:

court	first	enact	prosecution
basis	judge	legislation	civil
argument	kavanaugh	protect	liability
seems	written	presidents	officer
twofold	believes	criminal	mr

The percentage of negative words was then subtracted from the percentage of positive words. We then calculated the correlation between each of these dictionary-based congressional sentiment measures and the results from the Google API. These results are reported in the light grey bars of Figure 3. For both the overall and congressional sentiment, we also took the average of each dictionary score and calculated the same correlation. This measure is found in the "overall" columns. Ultimately, each correlation coefficient was statistically significant at the 0.001-level, but the correlations are noticeably higher for our measure of overall sentiment. For more details, please consult pages 15 – 17 of the main text.

S4 Vocal Pitch and Emotional Intensity

One of the central claims of this paper is that vocal pitch can be used as an indicator of emotional intensity. Since most in political science are not aware of this relationship, we dedicate this section to explaining the theoretical basis for our measure and introduce several validation exercises.

S4.1 The Circumplex Model of Affect and Vocal Pitch as a Measure of Emotional Activation

The present study draws heavily from James Russell's work (e.g., Russell 1980, 2003). More specifically, the circumplex model of affect suggests all affective states arise from two neurophysiological systems, one related to a pleasure-displeasure continuum (called "valence") and the other related to alertness (called "arousal" or "activation"). According to Russell (2003), at any given moment, one's emotional disposition is a single blend of these two dimensions with the horizontal axis (or valence) ranging from one extreme (e.g., unpleasant) to its opposite extreme (e.g., pleasant). Our study is mostly interested in the vertical (or arousal) dimension, which ranges from a deactivated emotional state, such as being sleepy, to an activated emotional state, ultimately culminating in "frenetic excitement" (Russell 2003, 148). In the context of legislative speech, we call this "emotional intensity."

As we explain in Section S4.3, vocal pitch is not the only way to measure arousal/intensity, but we think it is a reasonable approach. Indeed, in their review of emotional measurements Mauss and Robinson (2009) state:

> The assessment of vocal characteristics appears to be especially useful in understanding levels of emotional arousal, with higher levels of pitch and amplitude associated with higher levels of arousal (Table 1). By contrast, attempts to link emotional valence or discrete emotions to vocal characteristics have been met with mixed success at best, although more sophisticated methods may be capable of doing so in the future. Thus, we conclude that vocal characteristics are primarily reflective of the dimension of emotional arousal (225-226).

Generally speaking, when we are emotionally activated our muscles – including our vocal cords – naturally tighten. Posner, Russell and Peterson (2005) explain when sensory stimuli are present the reticular formation (RF) relay through the amygdalorecticular pathways (Koch and Ebert 1993; Rosen et al. 1991) increasing activity in the cerebral cortex (Heilman, Watson and Valenstein 2011; Jones 2003). This triggers a change in muscle tone and an increase in sweat production (Jones 2003) both of which are correlated with subjective ratings of emotional arousal (Lang et al. 1993). The resulting blood flow causes all muscles, including the vocal cords, to contract naturally, ultimately raising the pitch of one's voice.

Several scholars have found evidence that increased vocal pitch is associated with emotional activation. For example, Bachorowski and Owren (1995) randomly assigned positive ("Good Job") and negative ("Try Harder") feedback to subjects completing a word identification task. Subjects would then read some text aloud and answer some questions about their emotional state. Ultimately, the authors found that vocal pitch was higher when individuals reported increased levels of emotional intensity, leading them to conclude that "vocal pitch can be used to assess the level of emotional arousal currently experienced by the individual" (Mauss and Robinson 2009, 222). Pisanski, Nowak and Sorokowski (2016) conducted a similar experiment, but instead of randomly assigning feedback, they used the days leading up

to an oral exam to induce a similar effect in their subjects. Ultimately, they found the mean and minimum vocal pitch "increased significantly under stress" (236) and these changes in vocal pitch were significantly correlated with cortisol levels—a hormone linked to stress.

As we discuss in Section 6 of the main text, there is some concern MCs can strategically manipulate their vocal pitch to garner additional media coverage. However, this is unlikely since the automatic tightening of the vocal cords, like many physiological responses, is thought to occur largely below conscious awareness. Indeed, several studies have shown "verbal content" (i.e., the words spoken) tends to be at the "controllable end of the continuum, whereas the body and tone of voice may be classified as less controllable and more leaky channels" (Zuckerman and Driver 1985, 130). Moreover, when individuals attempt to control their vocal pitch this "emotional leakage" actually becomes more pronounced (Elkins et al. 2014). Whether it is a friend feigning laughter or a politician displaying certain emotions for strategic purposes, more work is required to convince others of false feelings. This naturally causes vocal pitch to increase in "deceptions which involve emotion" since such effort also tends to produce stress (Ekman et al. 1991, 133). Indeed, "deliberate attempts by liars at controlling expressive behaviors, such as attempts to control thoughts and feelings, can be the seeds of their own destruction" (DePaulo et al. 2003, 78).

Even though we show below vocal pitch is correlated with emotional intensity, vocal pitch is also "inversely proportional to vocal fold length" which is why men typically speak with a lower pitch than women (Puts, Gaulin and Verdolini 2006, 284). Given that, we encourage scholars to scale vocal pitch to standard deviations above or below each speaker's baseline (or mean) vocal pitch. Such standardization also helps "cancel out" any irregularities that may occur with the pitch extraction algorithm. Such a measure also helps isolate whether increases or decreases in vocal pitch influence an outcome of interest. Although we acknowledge the importance of better understanding the significance of raw pitch changes (e.g., Klofstad 2016), we are mostly interested in whether relative changes in emotional intensity – as measured by the speaker's vocal pitch – are predictive of news coverage. The standardize measure we introduce on page 13 of the main text is well suited to assess this relationship.

S4.2 Validation Exercises

Although we believe vocal pitch can be used to answer a variety of political science questions, we recognize that our measure is unfamiliar to many in the field. Given that, we have went to great lengths to validate vocal pitch as a measure of emotional intensity. For example, in Section S5.4 we show that Democrats and Republicans tend to speak with higher vocal pitch when talking about their respective party's owned issues. We also show that Democrats and Republicans who speak with more emotional intensity about party issues tend to vote with their party. This result is consistent with MCs naturally speaking at a higher vocal pitch when addressing issues they care about. In the sections that follow, we provide four additional validation exercises – all of which provide strong evidence that vocal pitch captures

Russell (2003)'s activation/intensity dimension.

S4.2.1 Validation #1: Data from Goudbeek and Scherer (2010)

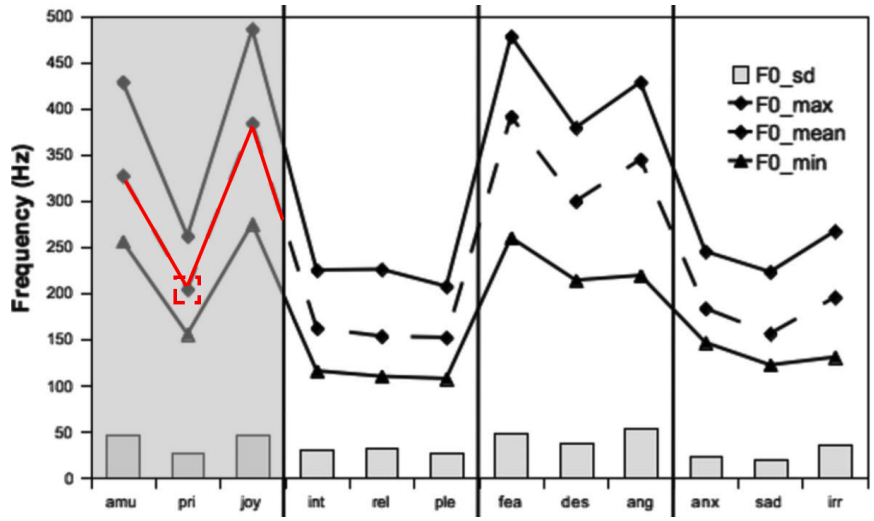
We first reprint relevant results from Goudbeek and Scherer (2010) who used the Geneva Multimodal Emotion Portrayals (GEMEP) corpus to understand how vocal characteristics are related to arousal and valance. Of the twelve emotions included in the corpus, six are positive and six are negative. Within each of the positive and negative categories, Goudbeek and Scherer (2010) identified three emotions as being "high intensity" and three emotions being "low intensity." This makes the GEMEP corpus well-suited to provide some additional validation for vocal pitch as a measure of emotional intensity, even though it utilizes trained actors. For those who are concerned about the use of actor portrayals, please consult Bänziger and Scherer (2007).

Generally, there are four arguments advanced supporting the use of actors for understanding the vocalization of emotion. First, acted expressions force researchers to provide precise definitions of the emotions under consideration rather than inferring the emotion post hoc. Second, actor portrayals are often the only way to study different emotions within the same individual. This is especially important for understanding emotional intensity, since an intense expression for one individual may not be an intense expression for another. Third, it is difficult to define the range of contexts in which natural audio should be obtained. It is even more difficult to fully conceptualize how those same settings influence the vocalization of the emotion itself. In this way, actor portrayals trade some external validity for greater internal validity. Finally, acted expressions can be measured in a number of different ways. For example, databases like GEMEP often include audio, video, and text, meaning multimodal studies of emotion are possible when using actors.

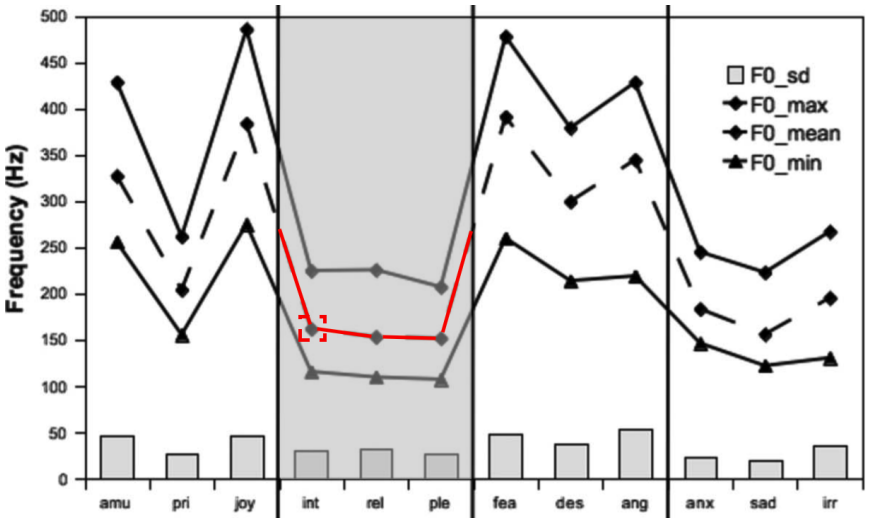
As a validation exercise, we reprint the second panel of Figure 1 from Goudbeek and Scherer (2010). As we show in Figure S2, their data demonstrate the relationship between vocal pitch and emotional intensity. In Panel A, we highlight high-intensity positive emotions, which include "Amusement," "Pride," and "Joy," in the grey box. The red line indicates the mean vocal pitch across these three emotions, and the red box identifies the lowest mean pitch (for "Pride"). In Panel B, we highlight low-intensity positive emotions, which include "Interest," "Relief," and "Pleasure," in the grey box. The red line indicates the mean vocal pitch across these three emotions, and the red box identifies the highest mean pitch (for "Interest"). Comparing across these two panels, it is clear that the mean vocal pitch for these high-intensity positive emotions is consistently higher than the mean vocal pitch for low-intensity positive emotions.

In Figure S3, we demonstrate a similar finding for the mean vocal pitch for negative emotions. In Panel A, we highlight the high-intensity negative emotions, which include "Fear," "Despair," and "Rage," in the grey box. The red line indicates the mean vocal pitch across these three emotions, and the red box identifies the lowest mean pitch (for "Despair"). In Panel B, we highlight the low-intensity negative emotions, which include

Figure S2: Reproduction of Figure 1, Panel 2 from Goudbeek and Scherer (2010) (Positive Emotions)



(a) High Intensity



(b) Low Intensity

Note: Reproduction of Figure 1, Panel 2 from Goudbeek and Scherer (2010). Additional grey shading and red highlights added to aid interpretation. In Panel A, red highlighting indicates mean vocal pitch for high intensity positive emotions, with a red box indicating the lowest mean vocal pitch. In Panel B, red highlighting indicates mean vocal pitch for low intensity positive emotions, with a red box indicating the highest mean vocal pitch.

"Anxiety," "Sadness," and "Irritation," in the grey box. The red line indicates the mean vocal pitch across these three emotions, and the red box identifies the highest mean pitch (for "Irritation"). As with positive emotions, we see that low-intensity negative emotions always have a lower vocal pitch than high-intensity negative emotions. Taken together, these previous results provide strong evidence that vocal pitch can reasonably measure the intensity of emotions.

S4.2.2 Validation #2: Data from Laukka (2004)

We next reprint results from Laukka (2004). Here, actors were asked to portray five base emotions – anger, disgust, fear, happiness, and sadness – at "high" and "low" intensity. Ultimately, Laukka (2004) also finds vocal pitch is correlated with emotional intensity which gives us additional confidence that our approach is reasonable. Similar to Section S4.2.1, we reprint the first panel of Laukka (2004)'s Figure 1. This can be found in Figure S4.

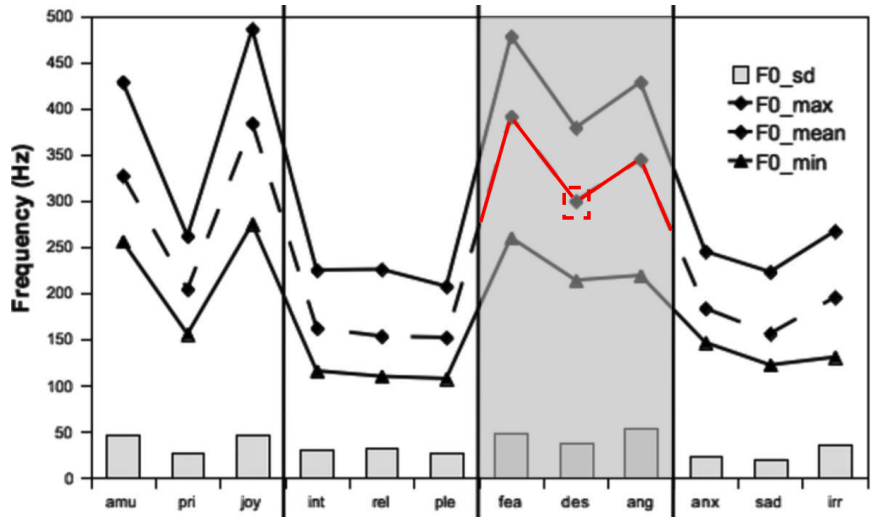
When actors are asked to display "high" and "low" intensity emotions it is clear that vocal pitch is significantly higher when emotions are portrayed with more intensity. More specifically, when Figure S4, Panel A is compared to Panel B it is clear that vocal pitch is substantially higher when actors are displaying emotions with "high intensity." Indeed, not only is the mean vocal pitch highlighted in Panel A always higher than the mean vocal pitch highlighted in Panel B, but none of the confidence intervals overlap. This demonstrates vocal pitch can be used to reasonably differentiate between more and less intense emotional expressions.

S4.2.3 Validation #3: New Analysis of Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS)

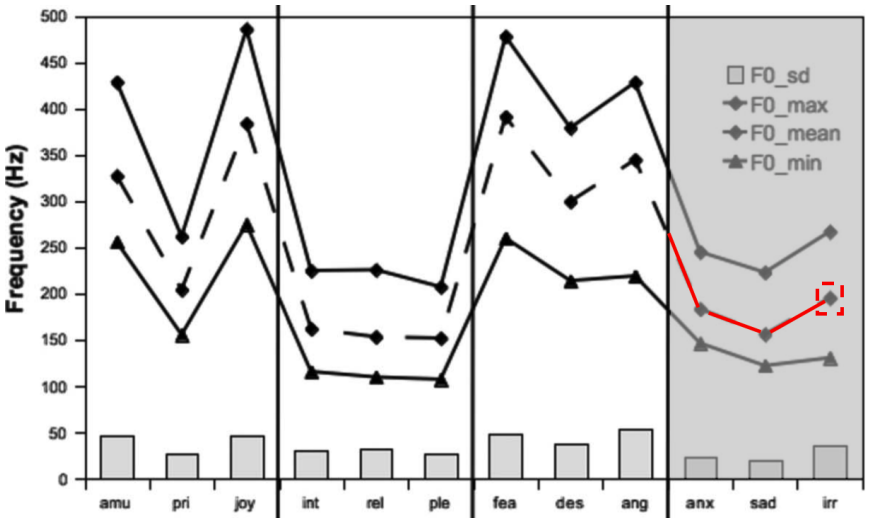
We conduct an analyze the Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) database in our third validation exercise. This corpus consists of 7,356 high-quality recordings of emotionally-neutral statements, spoken or sung with a range of emotions by 24 American male and female actors. Similar to Laukka (2004), in the speech data actors are asked to display 8 emotions – neutral, calm, happy, sad, angry, fearful, surprise, and disgust – with either "normal" or "strong" intensity. Nearly 300 independent coders then validated the portrayals by manually labeling the recordings for the valence and intensity of the emotional expression.

Our validation results can be found in Figure S5. Dark grey bars indicate the speaker was portraying an emotion with strong (or high) emotional intensity, whereas light grey bars indicate the speaker was portraying the same emotion with normal (or low) emotional intensity. In all plots, we used the default *Praat* settings. We show 95% confidence intervals for each estimate. Finally, we subdivided the data by male and female speakers since women tend to speak at a higher vocal pitch than men. Please see pages S5 and S20 for more details on gender-based differences in vocal pitch.

Figure S3: Reproduction of Figure 1, Panel 2 from Goudbeek and Scherer (2010) (Negative Emotions)



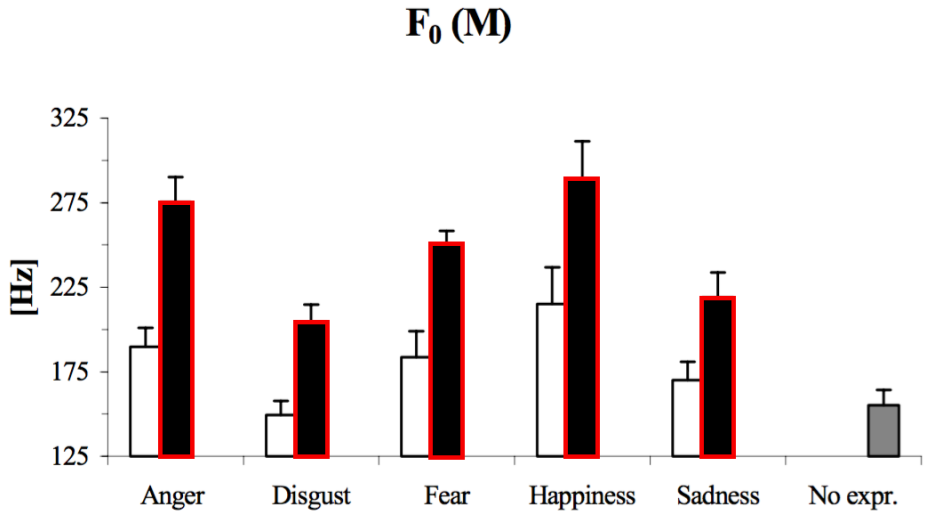
(a) High Intensity



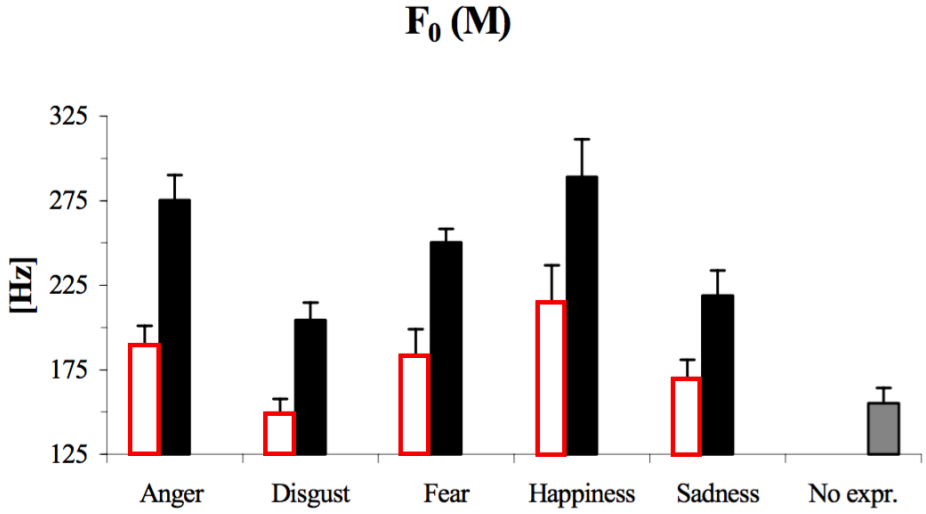
(b) Low Intensity

Note: Reproduction of Figure 1, Panel 2 from Goudbeek and Scherer (2010). Additional grey shading and red highlights added to aid interpretation. In Panel A, red highlighting indicates mean vocal pitch for high intensity negative emotions, with a red box indicating the lowest mean vocal pitch. In Panel B, red highlighting indicates mean vocal pitch for low intensity negative emotions, with a red box indicating the highest mean vocal pitch.

Figure S4: Reproduction of Figure 1, Panel 1 from Laukka (2004)



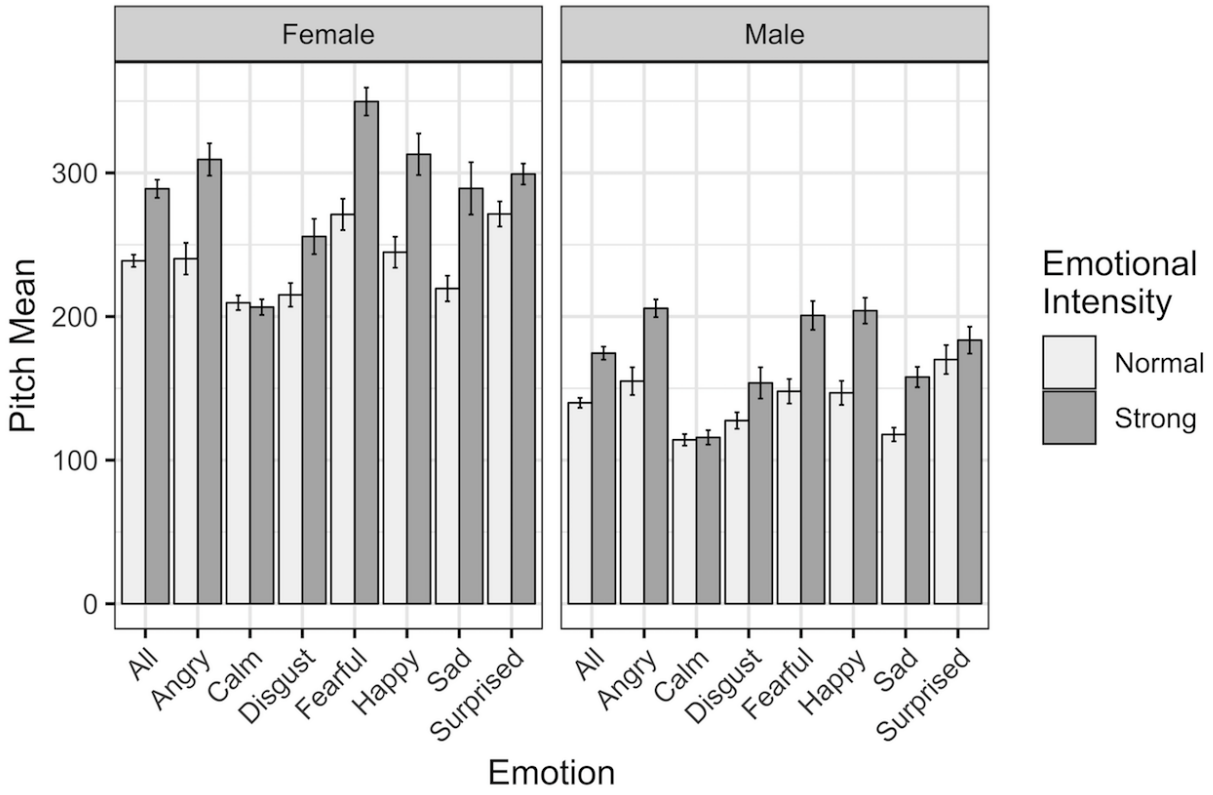
(a) High Intensity



(b) Low Intensity

Note: Reproduction of Figure 1, Panel 1 from Laukka (2004). Red boxes added to aid interpretation. In Panel A, red boxes indicate emotions expressed with high intensity. In Panel B, red boxes indicate emotions expressed with low intensity.

Figure S5: Emotions with "Strong" Intensity Delivered at a Higher Vocal Pitch Than Emotions with "Normal" Intensity (RAVDESS)



Note: Figure uses audio data from Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS). Light grey boxes indicate emotions displayed with "normal" intensity. Dark grey boxes indicate emotions displayed with "strong" intensity. Base emotions displayed on the x-axis. Red boxes highlight the trend across all emotions.

Our findings are similar for both men and women. First, when all emotional categories are combined (bars labeled "All" and highlighted in red), high-intensity emotions are delivered at a significantly higher vocal pitch. This is the most important result for our study, since it again suggests mean vocal pitch can be used to effectively discriminate between different levels of emotional intensity. Second, vocal pitch is significantly higher for high-intensity emotions in every category, except for "Calm." We think this is telling. Not only is it difficult to think of what "high intensity" calmness looks like, but Russell (2003) actually describes emotional activation as "frenetic excitement" (148). Given that, we contend calmness is synonymous with a less activated emotional state which is why it is not surprising that the speakers conveyed the high and low versions of this emotion using similar vocalizations.

S4.2.4 Validation #4: New Analysis of Giannakopoulos and Pikrakis (2014) Annotated Speaker Data

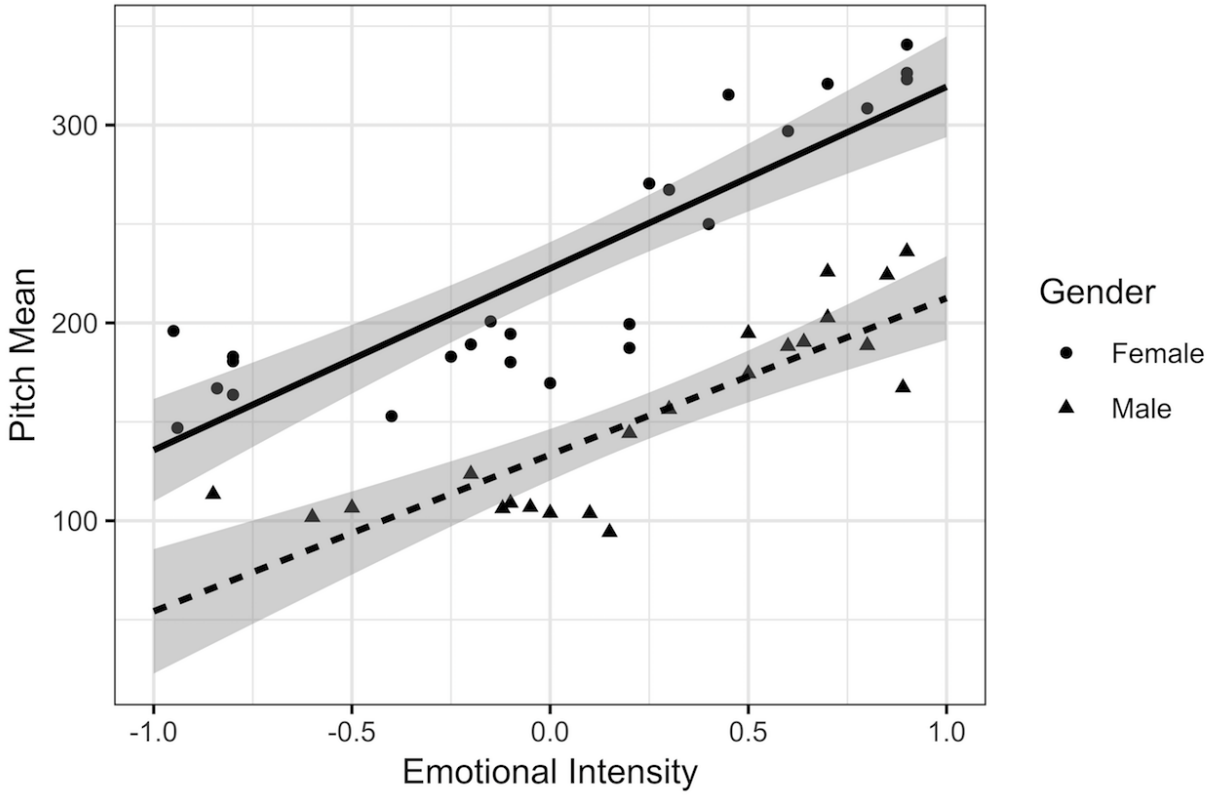
In our final validation exercise, we analyzed data provided by Giannakopoulos and Pikrakis (2014). This corpus consists of 47 audio clips collected from the Berlin Database of Emotional Speech. Each clip (spoken in German) was scored using a continuous scale ranging from low (-1) to high (1) emotional valence and activation. Unlike the RAVDESS data, there is little information about the degree to which the coding of valence and activation has been validated. For this reason, these data should only be used in conjunction with the other validation exercises we report above.

Figure S6 displays the results of our analysis. The x-axis represents the emotional intensity of the recordings, as coded by Giannakopoulos and Pikrakis (2014). The y-axis represents the mean vocal pitch of these recordings as extracted by *Praat*. Circles represent emotional portrayals from male actors, and triangles represent portrayals from female actors. We also present the result of simple linear models in which mean vocal pitch is regressed on the level of emotional intensity separately by speaker sex, with grey bands representing the respective 95% confidence intervals. Not only does this plot show a strong linear trend, but emotional intensity is correlated with the mean vocal pitch at the 0.84 level for men ($t = 8.66, df = 23, p < 0.001$) and 0.87 level for women ($t = 7.04, df = 20, p < 0.001$). As with our previous validation exercises, we again find higher mean vocal pitch is generally associated with more emotional intensity.

S4.3 Vocal Pitch: A Useful, but Not Exhaustive Measure

Although there are a number of ways to measure emotional intensity, we used vocal pitch in this study for three reasons. First, the present study wants to encourage the use of audio data in political science and we think vocal pitch is a good introductory measure. Not only are there well-established theoretical arguments for what vocal pitch captures, but it is relatively easy to estimate and extract. Just as text analysis has a place for dictionary methods, audio analysis should have a place for specific features – like vocal pitch. Indeed,

Figure S6: Emotions with "Strong" Intensity Delivered at a Higher Vocal Pitch Than Emotions with "Normal" Intensity (Giannakopoulos and Pikrakis)



Note: Figure uses data provided by Giannakopoulos and Pikrakis (2014). Solid and dashed lines represent simple linear regression lines for male and female speakers, respectively. Perceived emotional activation (or intensity) is shown on the x-axis. This variable ranges from (-1) deactivated to (1) activated. The speaker's raw vocal pitch is shown on the y-axis.

before we begin to utilize more advance methods, like supervised and unsupervised learning algorithms, we must first understand theoretical underpinning of specific features, like vocal pitch.

Second, vocal pitch also has a well-established literature *within* political science. Scholars like Casey Klofstad (e.g., Klofstad 2016) have been working on understanding the role vocal pitch has to play in social science research for quite some time. These scholars have made considerable strides in understanding vocal pitch using small-n studies, but we are the first to apply such techniques to large audio corpora. Just as "all quantitative models of language are wrong – but some are useful," (Grimmer and Stewart 2013, 269) we view vocal pitch in a similar light, especially when it is used at scale. With a large amount of audio data, we get a very good estimate of a speaker's baseline vocal pitch, which makes vocal pitch a much more useful measure because we have a good sense of what to expect.

Finally, "there is no globally best method for automated text analysis" (Grimmer and Stewart 2013, 270). We argue the same can be said for audio data. Indeed, much of the debate surrounding vocal pitch "can be resolved simply by acknowledging that there are different research questions and designs that imply different types of models will be useful" (Grimmer and Stewart 2013, 270). In this study, we are interested in whether certain types of floor speeches are more likely to appear on CNN, Fox News, and MSNBC. We think vocal pitch is a reasonable place to start this line of inquiry. Just as we hope this is not the last study on vocal pitch, we hope this is not the only audio variable scholars use in their own research. This is why the broader contribution of this study is not the use of vocal pitch, but showing an important way audio can be incorporated into political science research.

S5 Descriptive Statistics

S5.1 Description of Independent and Dependent Variables

As we explain in Section 4 of the main text, our independent and dependent variables come from a variety of sources. To help readers better understand the variables we used, we include the variable, a detailed description, and the source in Table S5.1. Generally speaking, most of our variables are calculated at the speech- and MC-level with the former representing 10 of the 19 variables we created for this study. Our focus on speech-level variables follows directly from the hypotheses we outlined in Section 3 of the main text. Although MC characteristics likely influence coverage, we are mostly interested in what makes certain floor speeches more or less likely to appear on CNN, Fox News, and MSNBC. Given that, a speech-level focus made the most sense for our study.

Table S3: Variable Descriptions

Variable	Description	Source	Level
Televised	A binary variable indicating whether the speech appeared on CNN, Fox News, and MSNBC. The variable is described on 18 of the main text. It was created using the *AudioNewser* which is described on pages 10 – 12 in the main text and pages S5 – S8 in the Supplemental Information.	*HouseLive* and *AudioNewser*	Speech
Total Coverage	A count variable capturing the total number of times the speech appeared on CNN, Fox News, and MSNBC. The variable is described on 18 of the main text. It was created using the *AudioNewser* which is described on pages 10 – 12 in the main text and pages S5 – S8 in the Supplemental Information.	*HouseLive* and *AudioNewser*	Speech
Total Minutes	A continuous variable capturing the total number of minutes the speech appeared on CNN, Fox News, and MSNBC. The variable is described on 18 of the main text. It was created using the *AudioNewser* which is described on pages 10 – 12 in the main text and pages S5 – S8 in the Supplemental Information.	*HouseLive* and *AudioNewser*	Speech
Total Viewers	A continuous variable capturing the total number of CNN, Fox News, MSNBC viewers for each speech. This was estimated using the total viewership (according to the 2014 Nielsen ratings) for each program in which the speech appeared. The variable is described on 18 of the main text. It was created using the *AudioNewser* which is described on pages 10 – 12 in the main text and pages S5 – S8 in the Supplemental Information.	*HouseLive* and *AudioNewser*	Speech
Vocal Pitch	Standardized vocal pitch measure which is described on page 13 of the main text and validated in Section S4 of the Supplemental Information. To account for any extreme values, we also re-estimated all the models excluding outliers in Section S7.4. These results are reported in Table S15.	*HouseLive* and *Praat*	Speech

Overall Sentiment	Standardized sentiment measure which is described on page 14 of the main text and validated in Section 4. (Please see Figure 3). To account for any extreme values, we also re-estimated all the models excluding outliers in Section S7.4. These results are reported in Table S15.	*HouseLive* and *Google API*	Speech
Congress Sentiment	Standardized congressional sentiment measure which is described on page 14 of the main text and validated in Section 4. (Please see Figure 3). To account for any extreme values, we also re-estimated all the models excluding outliers in Section S7.4. These results are reported in Table S15.	*HouseLive* and *Google API*	Speech
Republican	Binary variable which equals 1 if the speaker was a identified by *Voteview* as being a member of the Republican Party. The variable is described on page 17 of the main text.	*Voteview*	MC
|DW-Nominate|	Continuous variable which is the absolute difference between the speaker's DW-nominate score and the party median. The variable is described on page 17 of the main text.	*Voteview*	MC
Party Issue	Standardized measure capturing the degree to which speakers are referencing party issues. The measure is described on page 18 of the main text. Additional details are provided in Section S6 of the Supplemental Information. (Please see pages S42 – S42)	*HouseLive* and *STM Package*	Speech
Seniority	The total number of years served in the U.S. House of Representatives. The variable is described on page 17 of the main text.	*GovTrack*	MC
House Leader	Binary variable which equals 1 if the speaker was the current Speaker of the House, Majority/Minority Leader or Majority/Minority Whip. The variable is described on page 17 of the main text.	*House.gov*	MC
Committee Chair	Binary variable which equals 1 if the speaker was a committee chair. The variable is described on page 17 of the main text.	*Stewart and Woon*	MC

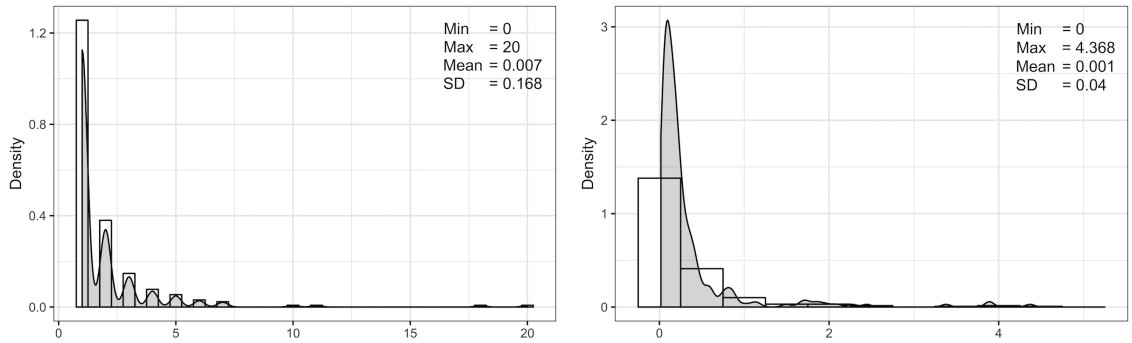
Male	Binary variable which equals 1 if the speaker was identified as a male by *GovTrack*. The variable is described on page 17 of the main text.	*GovTrack*	MC
White	Binary variable which equals 1 if the speaker was identified as African-American, Asian, or Hispanic/Latino by the House website. The variable is described on page 17 of the main text.	*House.gov*	MC
CQ-Bills	The number of CQ Bills being debated on the current legislative day as measured by CQ Weekly's "Bills to Watch." The variable is described on page 17 of the main text.	*CQ Weekly*	Day
One Minute	A binary variable indicating whether the speech was a one minute speech. The variable is described on page 17 of the main text.	*HouseLive*	Speech
Duration	A continuous variable capturing the length of the speech (in minutes). The variable is described on page 17 of the main text.	*HouseLive*	Speech
Election Year	A binary variable indicating whether the speech was delivered during an election year. The variable is described on page 17 of the main text.	*HouseLive*	Year

Given our reliance on speech-level measures, we were especially concerned about within MC dependence. We address this issue on page 18 and footnote 7 in the main text. Please also look at Section S7.6 in the Supplemental Information for additional model specifications that help address this issue. For any other concerns related to our variables or modeling choices, please consult the variable descriptions in Table S5.1. There we try our best to highlight the relevant sections of our manuscript for each of our independent and dependent variables.

S5.2 How Often Do Floor Speeches Appear On Television?

Floor speeches give MCs the ability to "send out a short, pointed message of their choice to a relatively large audience" (Morris 2001, 102). However, it is unclear the actual size of C-SPAN's audience. Table S4 shows not only are floor speeches aired on CNN, Fox News, and MSNBC, but when they do they have the potential to reach millions of viewers. Viewership was determined using 2014 Nielsen ratings. The average number of viewers for each show was then scaled to millions of viewers, with *The O'Reilly Factor* having the highest ratings at 2.667 million viewers. Even though this represents a small portion of the American public

Figure S7: Kernel Density and Histograms Plots For Total Coverage and Total Minutes



(a) Total Coverage (b) Total Minutes

Note: The kernel density and histogram plots are shown in grey and white, respectively. Both were created using the default settings in `R`. The minimum, maximum, mean and standard deviation are shown in the upper right hand corner.

and it is impossible to say whether viewers actually saw the floor speeches when they were aired, we use the average number of viewers to estimate a floor speech's potential reach.

We note three things in Table S4. First, CNN is much less likely to air floor speeches as compared to Fox News and MSNBC. This suggests CNN may view floor speeches as being less "newsworthy" as compared to their competitors. Whether this is grounded in specific broadcasting norms is unclear, but the results seem to suggest they view floor speeches differently. Second, the airing of floor speeches cannot be attributed to a single program. Regardless of whether one is talking about CNN, Fox News, or MSNBC, floor speeches are aired on a wide range of programs. Third, floor speeches also seem to make their way onto the most popular shows for each network, such as *Anderson Cooper 360*, *Hannity*, and *Hardball*. Not only do these shows have higher ratings than standard "newsroom" programs, but they are also aired and re-aired during primetime which gives these shows a considerable footprint.

To give a better sense of the overall distribution of our variable we created relevant density plots in Figures S7 and S8. Beginning with Panel A in Figure S7, we see the vast majority of floor speeches do not appear on cable news. Indeed, the mean number of tv appearances is 0.007 (see top right corner) suggesting the odds of a speech being repeatedly aired is low. The same can be said for the total number of minutes (see Panel B). Here, the maximum floor speech clip was 4.368 minutes which again suggests that floor speeches appear on CNN, Fox News, and MSNBC, but they are rarely shown in their entirety. With that said, we fully acknowledge this point on page 11 of the main text which begs the question – what is the value of delivering a floor speech?

Although we think this is a useful question, it is better to think of the perceived value which we describe on page 27 in the main text. MCs understand the odds of a speech being

Table S4: Which Cable News Programs Air Floor Speeches?

(a) CNN

Program	Floor Speeches	Viewers
Newsroom	13	0.467
Anderson Cooper	9	0.597
Out Front	8	0.498
New Day	6	0.317
Early Start	4	0.189
Situation Room	4	0.537
Saturday Morning	3	0.192
Around the World	2	0.390
State of the Union	2	0.518
Starting Point	2	0.192
Total	53	0.434

(b) FOX

Program	Floor Speeches	Viewers
America's Newsroom	30	1.264
Special Report	30	1.985
Hannity	27	1.499
Fox and Friends	22	1.049
The O'Reilly Factor	18	2.667
Shepard Smith	18	1.150
Happening Now	16	1.071
News Headquarters	16	1.264
Fox Report	9	0.403
Geraldo at Large	8	0.403
Total	214	1.303

(c) MSNBC

Program	Floor Speeches	Viewers
The Last Word	24	0.647
News Live	23	0.269
Hardball	20	0.760
Jansing & Co.	18	0.265
Rachel Maddow	16	0.843
The Ed Show	15	0.521
Countdown	13	0.354
Morning Joe	10	0.347
All In	8	0.639
Way Too Early	7	0.212
Total	175	0.439

Note: In each panel we show the total number of floor speeches aired by each network, as well as the total number of floor speeches aired by specific programs. These counts do not represent unique floor speeches. In the last column, we report the average number of viewers (in millions) for each program in 2014. This year was chosen because it was the last year in our data.

picked up by cable news networks is low, but the potential for publicity is still enticing. This is especially the case when compared to C-SPAN coverage which is often offered as the primary way by which floor speeches reach the American public. For example, Maltzman and Sigelman (1996) ask "Why would a member willingly forgo the opportunity to be beamed live into the homes of millions of Americans?" and then suggests MCs have an incentive to speak to a small C-SPAN audience because they tend to vote more than others.

We suspect most MCs place very little weight on C-SPAN coverage. Instead, they deliver floor speeches knowing (1) the videos can be posted online or distributed via social media and (2) there is a chance the national media will pick up their speeches. Here, we make an important distinction between "expected" and "perceived" value. The expected value of any speech is low since most speeches reach a small number of people, but the perceived value maybe higher given the few costs associated with trying to deliver a good speech.

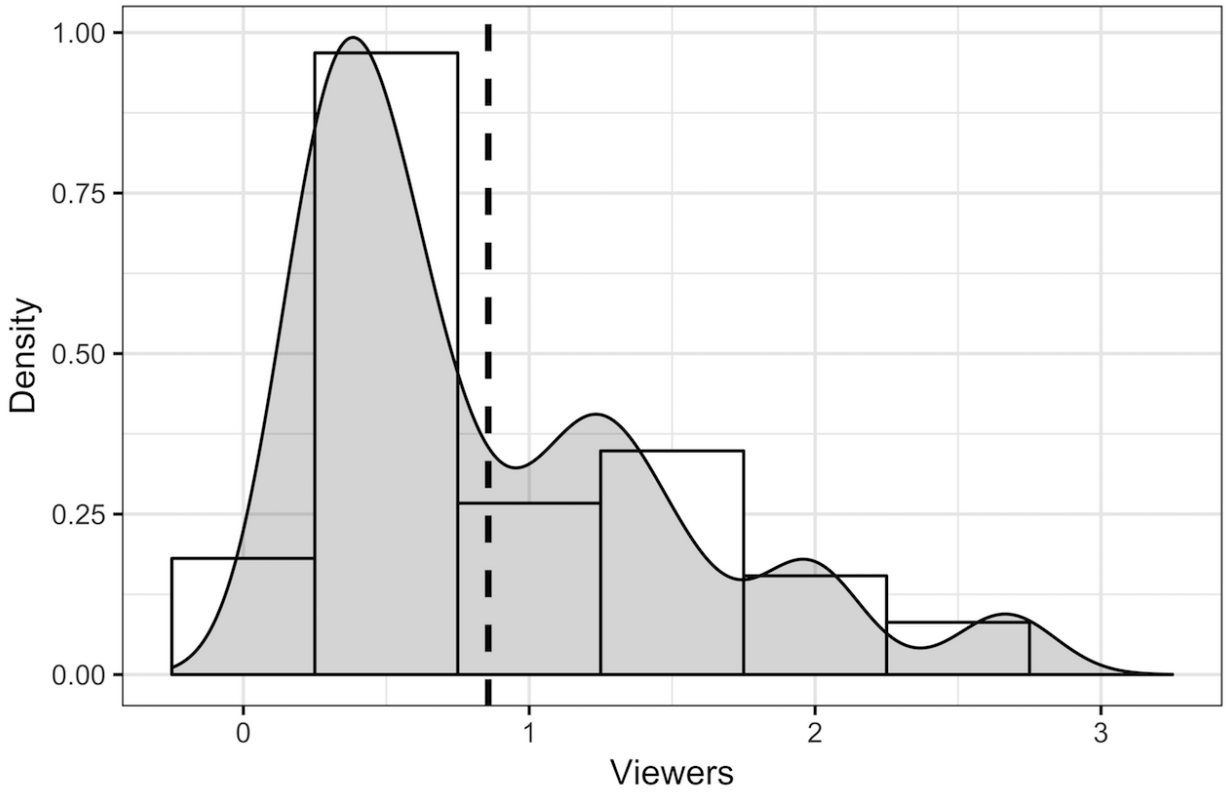
With that said, this discussion is very far removed from the present paper. This study does not aim to provide a comprehensive model of legislative speech. Instead, we are interested in better understanding some of the potential benefits floor speeches provide MCs and offer free air time on CNN, Fox News, and MSNBC as one of those benefits. Given that, this study should not be viewed as a treatise on all things related to legislative speech, but instead an exploration of one of the main ways floor speeches can reach the American public. Other mechanisms likely include social media posts and news bulletins, but we will leave these as avenues of future research.

We say cable news broadcasts are one of the main ways floor speeches reach the American public primarily because of the number of viewers MCs can reach when their speeches appear on CNN, Fox News, and MSNBC. This is shown in Figure S8. Here, the average floor speech reaches 856,796 viewers when it is aired on cable news. Although speeches appearing on Fox News received an average of 1,303,439 viewers, speeches appearing on MSNBC and CNN still received an average audience of 438,634 and 434,094, respectively. At first glance, this may suggest Republicans have a unique advantage, but Fox News aired Democratic speeches 55.61 percent of the time which is only 8.96 less than the MSNBC who broadcasted the highest percentage of Democratic speeches (64.57 percent). This suggests both parties' speeches receive comparable news coverage. Indeed, floor speeches delivered by Republicans reached, on average, 835,619 viewers, whereas floor speeches delivered by Democrats reached an average of 873,363 viewers.

S5.3 Most and Least Emotionally Activated Men and Women

We next consider which MCs deliver the most and least emotionally activated speeches. To create this table, we only consider MCs who delivered at least 5 speeches between 2009-2014. In Table S5, we list the 25 *most* (see Panel A) and least (see Panel B) activated male MCs. In the first and second column we report the average vocal pitch across all speeches in our data. In the second column we report the standard deviation. The third and fourth columns report the number of speeches and television appearances respectively. In the last three rows

Figure S8: How Many People View Floor Speeches When They Are Aired on Cable News Programs?



Note: This plot show the average number of viewers (in millions) for each floor speech using 2014 Nielsen Ratings. Vertical dashed line is the mean number of viewers. On average, when a speech was appeared on CNN, Fox News, or MSNBC it was viewed by 856,796 viewers.

Table S5: Most and Least Emotionally Activated Male Members of Congress

(a) *Most* Activated

	Pitch Mean	Pitch SD	N	TV Coverage
Name				
Capuano (D-MA)	196.49	15.48	36	0
Duffy (R-WI)	196.36	18.25	90	0
Herger (R-CA)	196.26	22.51	101	1
Hodes (D-NH)	192.11	5.36	7	0
Kennedy (D-RI)	191.37	29.84	45	18
Boustany (R-LA)	188.59	15.35	70	1
Frank (D-MA)	188.30	19.01	390	0
Paul (R-TX)	187.72	12.68	62	0
Marino (R-PA)	187.54	17.08	43	4
McClintock (R-CA)	186.90	19.18	252	9
Rush (D-IL)	186.51	22.04	72	0
Meeks (D-NY)	184.99	22.55	66	3
Wexler (D-FL)	184.87	38.75	5	0
Griffin (R-AR)	184.63	18.20	126	2
Buchanan (R-FL)	184.54	17.97	26	0
Mitchell (D-AZ)	184.49	18.22	9	0
Linder (R-GA)	184.30	18.29	25	0
Weiner (D-NY)	184.20	25.59	115	16
Dingell (D-MI)	184.07	26.37	110	0
Clarke (D-MI)	183.72	23.24	32	0
Kucinich (D-OH)	182.51	20.68	324	5
Amash (R-MI)	182.07	14.73	16	0
Landry (R-LA)	181.94	18.14	45	0
Pittenger (R-NC)	181.86	9.69	25	0
Woodall (R-GA)	181.50	16.80	357	0
Groups				
All	186.71	19.44	97.96	2.36
Democrats	186.97	22.26	100.92	3.50
Republicans	186.48	16.84	95.23	1.31

(b) *Least* Activated

	Pitch Mean	Pitch SD	N	TV Coverage
Name				
Snyder (D-AR)	122.76	14.17	20	0
DesJarlais (R-TN)	122.41	15.23	40	0
O'Rourke (D-TX)	122.22	9.61	28	0
Gallego (D-TX)	121.67	9.15	50	0
Pastor (D-AZ)	121.42	13.12	86	0
Gallegly (R-CA)	121.41	8.32	16	0
Petri (R-WI)	121.40	11.16	126	0
Rooney (R-FL)	121.39	11.64	49	0
McHugh (R-NY)	121.30	11.51	6	0
Latta (R-OH)	120.82	11.77	103	0
Tanner (D-TN)	120.71	14.49	26	0
Berry (D-AR)	120.58	11.81	7	0
Cummings (D-MD)	119.91	24.79	210	2
Holding (R-NC)	119.50	10.50	66	0
LaTourette (R-OH)	118.78	20.84	202	0
Rokita (R-IN)	118.20	16.24	67	0
Cooper (D-TN)	118.10	29.13	19	0
Benishek (R-MI)	116.55	8.45	63	0
Costello (D-IL)	116.40	11.57	56	0
Lankford (R-OK)	116.26	14.97	125	1
McCotter (R-MI)	114.90	12.36	91	1
Latham (R-IA)	114.49	15.13	169	0
Ruppersberger (D-MD)	112.99	15.33	60	0
Veasey (D-TX)	112.37	11.97	62	0
Rahall (D-WV)	110.47	14.52	101	0
Groups				
All	118.68	13.91	73.92	0.16
Democrats	118.30	14.97	60.42	0.17
Republicans	119.03	12.93	86.38	0.15

Note: Measurements of vocal pitch are in Hertz (Hz). The columns labeled "N" and "TV Coverage" report the total number of speeches and television appearance respectively. The 25 *most* (see Panel A) and *least* activated (see Panel B) male MCs were determined using their average vocal pitch across all speeches in our data. Here, we only included MCs who delivered at least 5 speeches. Column averages for Democrats and Republicans can be found in the "Groups" section.

we report column averages, including the mean vocal pitch for Democrats and Republicans.

Table S5 shows there is substantial variation in which male MCs are generally the most and least activated when speaking. For example, 11 of the 25 most activated male MCs (or 44 percent) were Democrats which is essentially the same as the overall party breakdown within our data where 44.20 percent of all male speeches were delivered by Democrats. We find nearly the same proportion in Panel B where 12 out of the 25 least activated male MCs (or 48 percent) were Democrats, suggesting neither party is especially intense when speaking. Importantly, we find that the congressmen who speak with the greatest intensity tend to appear more often on CNN, Fox News, and MSNBC. In fact, there is a statistically significant difference between the average number of television appearances for the 25 most (Table S5a) and least (Table S5b) activated men ($t = 2.23$, $df = 48$, $p \leq 0.04$). This suggests

Table S6: Most and Least Emotionally Activated Female Members of Congress

(a) *Most* Activated

	Pitch Mean	Pitch SD	N	TV Coverage
Name				
Sanchez (D-CA)	261.09	34.44	51	0
Blackburn (R-TN)	256.05	29.49	282	3
Markey (D-CO)	254.86	21.10	12	0
DeGette (D-CO)	248.63	27.91	48	0
Sutton (D-OH)	248.10	31.47	133	0
Castor (D-FL)	242.81	24.91	108	0
Halvorson (D-IL)	242.29	23.88	34	0
Wilson (D-FL)	238.63	37.02	30	0
Brownley (D-CA)	237.24	25.87	36	0
Shea (D-NH)	232.47	15.18	57	0
Kirkpatrick (D-AZ)	232.47	46.07	87	1
Baldwin (D-WI)	231.77	15.37	47	0
Kuster (D-NH)	230.42	21.73	43	0
Fallin (R-OK)	227.64	22.76	24	0
McCollum (D-MN)	227.23	25.49	66	0
Waters (D-CA)	226.30	17.34	208	1
Hayworth (R-NY)	225.17	30.13	31	0
Sinema (D-AZ)	224.57	15.00	23	0
Noem (R-SD)	223.07	49.81	51	0
Hartzler (R-MO)	222.78	19.37	106	0
Ellmers (R-NC)	220.09	17.60	94	2
Black (R-TN)	219.32	16.34	118	0
Schakowsky (D-IL)	219.19	25.08	235	2
Bachmann (R-MN)	218.29	27.96	297	7
Bonamici (D-OR)	218.04	20.44	57	0
Groups				
All	233.14	25.67	91.12	0.64
Democrats	236.24	25.19	75.00	0.24
Republicans	226.55	26.68	125.38	1.50

(b) *Least* Activated

	Pitch Mean	Pitch SD	N	TV Coverage
Name				
Jenkins (R-KS)	190.79	16.43	100	0
Brown (R-FL)	190.23	24.92	56	0
Wasserman (D-FL)	190.14	27.22	158	4
Emerson (R-MO)	190.05	25.37	60	0
Capito (R-WV)	189.98	17.86	168	0
Bean (D-IL)	189.21	17.85	28	0
Kosmas (D-FL)	188.92	10.94	16	0
Matsui (D-CA)	188.63	11.63	124	0
Berkley (D-NV)	187.90	17.71	132	0
Moore (D-WI)	187.83	20.38	185	1
Fudge (D-OH)	187.53	21.18	212	0
Clarke (D-NY)	186.84	15.56	78	0
Gabbard (D-HI)	186.62	8.08	53	0
DelBene (D-WA)	185.65	19.42	18	0
Davis (D-CA)	184.23	34.94	176	3
Beatty (D-OH)	183.62	15.68	48	0
Brooks (R-IN)	182.65	14.10	34	0
Tsongas (D-MA)	178.32	11.12	72	0
Capps (D-CA)	177.17	14.04	294	0
Biggert (R-IL)	175.33	14.80	213	0
Chu (D-CA)	173.59	13.08	198	0
Miller (R-MI)	170.73	31.54	176	5
Esty (D-CT)	169.45	13.95	36	0
Granger (R-TX)	156.47	8.95	22	0
Kelly (D-IL)	143.45	5.42	24	0
Groups				
All	181.41	17.29	107.24	0.52
Democrats	181.71	16.36	108.94	0.47
Republicans	180.78	19.25	103.63	0.625

Note: Measurements of vocal pitch are in Hertz (Hz). The columns labeled "N" and "TV Coverage" report the total number of speeches and television appearance respectively. The 25 *most* (see Panel A) and *least* activated (see Panel B) female MCs were determined using their average vocal pitch across all speeches in our data. Here, we only included MCs who delivered at least 5 speeches. Column averages for Democrats and Republicans can be found in the "Groups" section.

that the male legislators who are especially emotionally intense also tend to garner the most cable news coverage.

Undoubtedly, Patrick Kennedy (D-RI) and Anthony Weiner (D-NY) received considerably more coverage as compared to any of the other male MCs in either Table S5a or Table S5b. However, even when "TV Coverage" is converted to a binary variable a statistically significant difference is still found ($t = 2.03$, $df = 48$, $p \leq 0.05$). When a χ^2 test is used to compare the proportion of the most and least activated male MCs who appeared on television ($\chi^2 = 3.95$, $df = 1$, $p \leq 0.05$), a similar result is found which provides additional support for our first hypothesis.

In Table S6 we list the 25 *most* (see Panel A) and *least* (see Panel B) activated female MCs. Similar to their male counterparts, 17 of the 25 most activated female MCs (or 68

percent) were Democrats which is essentially the same as the overall party breakdown within our data where 69.54 percent of all female speeches were delivered by Democrats. We find the same proportion in Panel B where 17 out of the 25 least activated female MCs (or 68 percent) were Democrats, suggesting neither party is especially intense when speaking. Even though we do not find the most activated female MCs are significantly more likely to appear on television ($t = 0.29$, $df = 48$, $p \leq 0.78$), the direction of the relationship is the same as their male counterparts. Again, this is consistent with our first hypothesis.

S5.4 Most and Least Emotionally Activated Democrats and Republicans

We also determined whether party members tended to be more emotionally intense when speaking about party issues. To compute the "party speech" variable, we first calculated the average proportion of a speech dedicated to party issues based on the closed captioning of the speech. These issues were identified using the STM outlined in Section S6 and the issues listed in Table S8. If an MC referenced party issues more than we would expect given the mean, we said he/she was delivering a party speech (1). Otherwise, it was not considered a party speech (0). For all MCs, we then calculated their mean vocal pitch when they were and were not giving a party speech. The difference between these indicates the extent to which MCs were emotionally intense when speaking about their party's issues.

Using DW-NOMINATE scores, we then calculated each legislator's distance from the party median since we expect legislators closest to the party median are more likely to speak with emotional intensity about party issues since they are aligned with their own policy preferences. For example, Representative Neal (D-AL) has a much more conservative DW-NOMINATE score (0.06) than the Democratic median (-0.40). Since he votes less frequently with his Democratic colleagues, we expect he would be less excited to speak about his party's issues which should be reflected in a decrease in his vocal pitch.

Table S7 lists the 25 most- and least-activated MCs when talking about party issues. In accordance with expectations, we find that the 25 legislators most activated when talking about party issues had an average DW-NOMINATE distance of 0.11 from their party median, compared to a distance of 0.19 for the 25 least activated legislators. Although this difference is slight, it is still statistically significant at the 0.05-level ($t = 3.24$, $df = 48$, $p < 0.01$). This general result holds for both Democrats ($t = 2.43$, $df = 30$, $p < 0.03$) and Republicans ($t = 1.96$, $df = 16$, $p < 0.07$), even though the latter difference is only statistically significant at the 0.07-level.

We cannot say definitively that party members are the *only* group that speaks passionately about Democratic- and Republican-owned issues, nor can we say that only those members closest to their party's median voting behavior are emotionally intense on these issues. Still, these results provide another piece of predictive validity for our use of vocal pitch as an indicator of emotional intensity.

Table S7: Members of Congress Who Speak with Emotional Intensity About Party Issues Tend to Vote More With Their Party

(a) *Most* Activated

	Party Issue	No Party Issue	Pitch Diff.	|DW Diff.|
Name				
Sánchez (D-CA)	280.45	247.53	32.92	0.05
Carney (D-PA)	166.85	136.27	30.58	0.29
Wexler (D-FL)	202.62	173.03	29.59	0.00
Tauscher (D-CA)	220.03	192.17	27.86	0.10
Herrera (R-WA)	236.86	209.76	27.10	0.00
Meek (D-FL)	168.63	142.01	26.62	0.08
Holden (D-PA)	145.89	119.50	26.39	0.17
Napolitano (D-CA)	207.24	182.36	24.87	0.08
Roybal-Allard (D-CA)	209.44	188.53	20.91	0.01
Tiberi (R-OH)	152.14	131.24	20.89	0.08
Scott (R-GA)	165.65	145.42	20.23	0.04
Davis (D-IL)	137.95	117.80	20.15	0.06
Owens (D-NY)	173.91	153.82	20.09	0.23
Bonner (R-AL)	179.73	160.96	18.77	0.26
Schauer (D-MI)	151.64	133.03	18.61	0.06
Thompson (D-MS)	154.29	136.10	18.19	0.04
Bono (R-CA)	220.32	202.32	18.00	0.03
Wasserman (D-FL)	199.16	182.19	16.97	0.03
Amash (R-MI)	194.68	177.87	16.81	0.25
Myrick (R-NC)	225.12	208.92	16.20	0.04
Dingell (D-MI)	193.83	177.81	16.03	0.01
Kilpatrick (D-MI)	212.90	196.92	15.97	0.06
Rodriguez (D-TX)	150.02	134.16	15.86	0.11
Shuler (D-NC)	153.84	138.05	15.79	0.35
Lee (R-NY)	157.19	141.47	15.72	0.19
Groups				
All	186.41	165.17	21.24	0.11
Democrats	184.04	161.84	22.20	0.10
Republicans	191.46	172.24	19.21	0.11

(b) *Least* Activated

	Party Issue	No Party Issue	Pitch Diff.	|DW Diff.|
Name				
Amodei (R-NV)	133.02	160.13	-27.11	0.10
Murtha (D-PA)	147.66	174.50	-26.85	0.18
Green (D-TX)	155.33	181.73	-26.40	0.09
Cramer (R-ND)	153.89	175.37	-21.48	0.27
LoBiondo (R-NJ)	154.23	174.76	-20.54	0.25
Space (D-OH)	147.24	166.05	-18.81	0.25
Pomeroy (D-ND)	155.89	172.92	-17.02	0.18
Clarke (D-MI)	176.30	193.27	-16.97	0.08
Melancon (D-LA)	125.94	142.88	-16.94	0.24
Giffords (D-AZ)	204.09	220.54	-16.45	0.31
McHugh (R-NY)	107.96	123.97	-16.02	0.38
Cooper (D-TN)	106.02	121.33	-15.31	0.13
DelBene (D-WA)	181.40	196.69	-15.29	0.14
Adler (D-NJ)	141.06	156.33	-15.27	0.27
Heck (R-NV)	147.91	163.13	-15.23	0.08
Noem (R-SD)	212.15	227.20	-15.05	0.23
Cook (R-CA)	164.12	178.99	-14.87	0.12
Baird (D-WA)	132.08	146.79	-14.71	0.13
Moore (D-KS)	124.83	139.40	-14.58	0.18
Kelly (R-PA)	168.05	182.37	-14.32	0.29
Markey (D-MA)	154.66	168.85	-14.19	0.12
Gibson (R-NY)	141.05	155.21	-14.15	0.21
Miller (R-CA)	131.68	145.16	-13.48	0.13
Halvorson (D-IL)	237.54	251.01	-13.47	0.12
Peters (D-CA)	129.57	142.87	-13.29	0.25
Groups				
All	153.35	170.46	-17.11	0.19
Democrats	154.64	171.68	-17.04	0.18
Republicans	151.41	168.63	-17.22	0.20

Note: Measurements of vocal pitch are in Hertz (Hz). To validate vocal pitch as a measure of emotional intensity, we created a dummy variable which equals 1 when a MC's speech contained more than the average number of party references. We called these "Party Speeches." In the first column, we restricted our data to party speeches. In the second column, we restricted our data to speeches which contained less than the average number of party references. The "Pitch Difference" column (abbreviated "Pitch Diff.") is the difference between these two columns. The 25 *most* (see Panel A) and *least* activated (see Panel B) MCs had the highest and lowest "Pitch Difference," respectively. The absolute difference between the MC's DW-Nominate score and the median DW-Nominate score for the MC's party is included in the "DW Difference" (abbreviated "DW Diff.") column. Higher values imply the MC's ideology was further away from the party median. Column averages for Democrats and Republicans can be found in the "Groups" section.

S6 Controlling for Issue Attention

As explained on 18 in the main text, we control for a speech's issue content using a variant of the "party speech" variable described on page S39. More specifically, we use a Structural Topic Model (STM) to determine the proportion of the speech dedicated to party issues using the topics highlighted in Table S8. In this section, we give more details about the LDA-based topic model we used and the topics we ultimately selected as Democratic and Republican issues.

S6.1 Do Members of Congress Talk about Party Issues?

Unlike standard LDA models, the STM model gives researchers the ability to incorporate metadata, such as when a floor speech was given or whether the floor speech was delivered by a liberal or conservative member of Congress, in order to improve the estimated LDA (Roberts et al. 2013; Roberts, Stewart and Tingley 2014; Roberts et al. 2014). Similar to LDA, each document is modeled as a mixture of K topics. Unlike standard LDA, in STM, topic proportions (θ) can be correlated and the likelihood of those topics can be influenced by some set of covariates X which are modeled using a logistic normal regression. For each word (w) within a given document, a topic (z) is drawn from the document-specific distribution and conditioned on that topic. The word itself is chosen from a multinomial distribution which is parameterized by β and represents the deviations from the baseline word frequencies (m) in log space. This distribution can also vary by some set of covariates U. These additional covariates "structure" the topic model's prior distributions, allowing researchers to use their substantive knowledge of the corpus to facilitate topic identification (for additional details, see Roberts et al. 2014, 4).

To use an STM, the researcher has to make a number of choices, most notably the covariates used in order to fit the topic model and the number of topics. For this study, we included two covariates: the speaker's ideology and the date of the speech. The former was measured using DW-Nominate scores which range from -1 ("liberal") to 1 ("conservative"). These scores are calibrated using floor votes and have been consistently used as measures of congressional ideology (Poole and Rosenthal 2001). We assume that representatives who are on the same side of the ideological spectrum are likely to speak about similar issues, meaning the words used in those speeches are likely to be clustered. We also included the date of the speech as a covariate. This variable was measured in days since the first date in the data set – January 1, 2009. We assume that each legislative day is restricted to a handful of topics, meaning representatives are likely to deliver similar speeches on the same day. If this is the case, then one would expect words appearing on the same day are more likely to be associated with one another.

Unfortunately, in both instances, the exact relationship between each covariate and the topics being discussed is unknown. Given that, we fitted a b-spline (or basis spline) to both the speaker's ideology and the date of the speech. These smoothed covariates were the ones

ultimately used to estimate the STM. Unlike other smoothing functions that do not allow the curve to change locally without causing changes to the full length of the curve, a b-spline is fairly flexible, allowing both local cusps and additional points to be added without increasing the degree of the curve. For these reasons, the b-spline is considered to be "more numerically stable than the cubic spline" (Keele 2008, 59). It is unclear whether smoothing is necessary to fit the model, but b-splines have been used in other STM applications (e.g., Roberts, Stewart and Airoldi 2016), making this choice consistent with previous work.

Similar to other unsupervised models, the most important decision the researcher has to make is the number of topics (k). Roberts, Stewart and Tingley (2014) provide two measures that help researchers with this choice: *semantic coherence* and *exclusivity*. A topic is coherent when the average of pairwise word similarities formed by the topic's top words is high. The average exclusivity of a topic is the degree to which the topic's words appear in that particular topic to the exclusion of others. A "good" topic is one that maximizes both measures, meaning the topic has words that are similar to one another and are unique to that topic. Using these two measures as guidelines, we estimated nineteen structural topic models, with k ranging from 10 to 100, increasing in increments of 5. Ultimately, $k = 30$ and $k = 35$ yielded very similar results. To break the tie, we used the held-out likelihood which estimates the probability of words appearing within a document when those words have been removed prior to estimation (Blei, Ng and Jordan 2003). When this was done, a 30-topic model performed slightly better which is why we ultimately chose $k = 30$.

Table S8 shows the top 5 words appearing in each topic, as well as the average proportion of each speech dedicated to that topic. The labels were added post-hoc and hopefully help the reader understand the substantive importance of each topic. "Democratic issues" fell into three general categories: (1) social welfare, (2) land management/transportation, and (3) civil rights. These includes topic numbers 1, 2, 6, 13, 19, 22, 25, 28, and 29. For Republicans, we identified (1) defense, (2) immigration, and (3) tax/budget policy as "Republican issues." These include topic numbers 5, 8, 9, 10, 12, 15, 23, 26, and 30. Although there is no universally-accepted definition of which issues are owned by the Democratic and Republican parties, we feel that these selections are in keeping with general perceptions of the parties' issue ownership. Moreover, we find Democrats and Republicans dedicate a larger proportion of their speeches to topics owned by their parties ($t = 23.32, df = 74148, p < 0.001$ and $t = 30.46, df = 74148, p < 0.001$, respectively).

Looking to the column labeled "proportion" in Table S8 we can also gain some insights into the topics that generally receive the most attention on the House floor. For example, the second "discursive" topic was talked about the most, representing, on average, 8% of each speech. This is not too surprising since this topic seems to capture common discursive elements, like "think" and "know." Of the Republican issues, the topic labeled "spending cut" received the most attention, representing, on average, 4% of each speech. For Democratic issues, "health care" receives a similar amount of attention which is consistent with the issues outlined by Petrocik (1996) as being traditionally owned by the Democratic party.

With that said, these proportions should not be viewed in absolute terms. In the vast

Table S8: Democratic and Republican Issues Identified by the Structural Topic Model (STM) Outlined in Section S6

Topic	Word 1	Word 2	Word 3	Word 4	Word 5	Label	Proportion
1	court	case	justic	judg	law	law 1	0.01
2	right	peopl	constitut	american	freedom	rights	0.03
3	colleagu	support	today	like	new	collegiality 1	0.05
4	work	make	need	peopl	can	values	0.07
5	war	militari	afghanistan	forc	defens	middle east 1	0.02
6	school	educ	student	colleg	communiti	education	0.02
7	republican	american	democrat	will	pass	party	0.04
8	busi	small	regul	cost	will	business	0.02
9	budget	spend	cut	year	debt	spending cut	0.04
10	secur	nation	inform	protect	agenc	security	0.03
11	energi	oil	gas	will	price	energy	0.02
12	state	unit	texa	border	come	immigration 2	0.02
13	care	health	insur	will	cost	health care	0.04
14	women	children	famili	live	life	children	0.03
15	nuclear	israel	iran	world	peac	middle east 2	0.02
16	job	economi	creat	american	econom	jobs	0.04
17	peopl	get	thing	talk	got	discursive 1	0.05
18	honor	year	great	serv	first	collegiality 2	0.04
19	transport	build	new	system	air	transportation	0.02
20	financi	credit	loan	bank	street	financial	0.02
21	will	side	pass	floor	debat	procedural	0.05
22	water	land	area	communiti	nation	land	0.03
23	law	immigr	enforc	victim	crime	immigration	0.02
24	say	think	know	want	one	discursive 2	0.08
25	fund	program	million	provid	billion	welfare 1	0.04
26	tax	govern	pay	feder	american	tax	0.03
27	administr	quot	report	obama	public	administration	0.03
28	act	requir	author	law	provis	law 2	0.05
29	famili	food	benefit	million	cut	welfare 2	0.02
30	servic	veteran	nation	serv	support	veterans	0.03

Note: Blue and Red topics were flagged as Democratic and Republican issues, respectively. Top-5 words and labels from the ($k = 30$) STM outlined in Section S6 also included. The labels are not returned by the software. They were added after reviewing the top-5 words and other related output.

Table S9: Controlling for "Health Care" References

Table Number	Model Number	Vocal Pitch Coefficient	Overall Sentiment Coefficient	Congress Sentiment Coefficient	Health Care Coefficient
1	2	0.590**	0.205**	0.187†	0.186**
1	4	0.843**	0.304**	0.129	0.184**
1	6	0.271**	0.087**	0.077†	0.079**
1	8	1.232**	0.466**	0.320†	0.326**

Note: In the first and second columns, we list the table and model numbers, respectively. In the third column, we list the independent variable. The fourth column shows the results from the main text. In these models, we control for a speech's issue content using the proportion of the speech dedicated to party issues (see "Party Issue Coefficient"). In the fifth column, we replace this variable with with Table S8, Topic 13. This new variable ranges from 0 to 1 with 1 implying 100 percent of the speech was dedicated to health care. In these models, ever other variable is the same (see "Health Care Coefficient"). Levels of significance (with standard errors clustered around each MC) are reported as follows: $^{\dagger}p < .1$; $^{*}p < .05$; $^{**}p < .01$.

majority of topic models, the proportion of a speech dedicated to a single topic is incredibly low with topics capturing 10% of the speech being at the higher end of the spectrum. That is why we encourage scholars to think about topic proportions in relative terms, meaning they can be used to help understand what topics are more or less popular, but it is difficult to use them to say one topic is discussed 2% more than another. To do so, one would have to make strong assumptions about the lexical boundaries of each topic which most scholars are unwilling to do. The fact that we find partisans tend to speak significantly more about their party's topics (see page S42) gives us some confidence our topics are reasonable enough, especially for a control variable.

S6.2 Do Members of Congress Talk about Health Care?

As explained on 18 in the main text and the introduction to Section S6 of the Supplemental Information, we control for a speech's issue content using the proportion of the speech dedicated to party issues. However, we also understand that some issues are generally more newsworthy. For example, in 2009 and 2014 health care received a considerable amount of coverage on CNN, Fox News, and MSNBC which is why we wanted to conduct a separate analysis where we included this topic as an additional control.

To determine whether the discussion of health care had any effect on the models we report in the main text, we re-estimated all models including Table S8, Topic 13 which we labeled as "Health Care" as an additional control. In the first two columns of Table S9, we report the table and model number from the main text. For example, in the first row we use the model reported in Table 1, Model 2, but now we include "Health Care" as an additional control. In the columns 3–6, we report the coefficients for vocal pitch, overall sentiment, and congressional sentiment. These serve as our main independent variables. In the last column,

we report the coefficients for "Health Care" which is a significant predictor of cable news coverage suggesting our concerns have some face validity.

However, regardless of the dependent variable, the results remain the same. Both coefficients and levels of significance are essentially the same, especially for vocal pitch which is still a strong and positive predictor of cable news coverage. This suggests that discussions of health care cannot explain the results we report in the main text. Indeed, the main effects for overall and congressional sentiment still hold even with an additional control for the proportion of the speech dedicated to "Health Care." This underlines the robustness of our results.

S7 Alternative Model Specifications

S7.1 Network Specific Models

Table S10 reports network specific models. In Panel A, we report the results from the main text. In Panel B, we restrict our data to only include CNN broadcasts. We do the same for Fox News and MSNBC in Panels C and D respectively. In the first two columns of each panel, we report the corresponding table and model number from the main text. For example, in the first row of the table reported in Panel B we use the model reported in Table 1, Model 1 but now we only consider CNN broadcasts. In columns 3–6, we report the coefficients for vocal pitch, overall sentiment, and congressional sentiment. These serve as our main independent variables. Here, we are interested in whether the coefficients change when the data is restricted to only include CNN, Fox News, or MSNBC. For the most part, we find our results are the same across all three networks.

We begin with Panel B. Here, we only consider (1) whether a floor speech appeared on television, (2) the number of times the floor speech was televised, and (4) the estimated number of total viewers (in millions) for CNN broadcasts. Unfortunately, the model did not converge when the dependent variable was (3) the total air time (in minutes). Regardless of the model, `Vocal Pitch` remains a strong and statistically significant predictor, even though `Overall Sentiment` is no longer statistically significant at the 0.05-level. With that said, `Congress Sentiment` is also in the predicted direction and statistically significant when Table 1, Model 2 is re-estimated only using CNN broadcasts.

Similar results are found for Fox News (see Panel C), where `Vocal Pitch` is a strong and statistically significant predictor regardless of the dependent variable. The same can be said for `Overall Sentiment`, even though `Congress Sentiment` fails to reach traditional levels of statistical significance in any of the re-estimated models. With that said, `Congress Sentiment` is a statistically significant predictor when Table 1, Models 1 and 3 are re-estimated only using MSNBC broadcasts. Although consistent with our third hypothesis, we find must stronger support for our first and second hypotheses in Panel D. Indeed, here we find both `Vocal Pitch` and `Overall Sentiment` are strong and statistically significant

Table S10: CNN, Fox News, and MSNBC Coverage of Verbal and Non-Verbal Emotional Expressions on the Floor of the U.S. House of Representatives (Network Specific Models)

(a) All

Table Numb.	Model Numb.	Vocal Pitch Coeff.	Overall Sent. Coeff.	Cong. Sent. Coeff.
1	1	0.617**	0.237**	0.181†
1	2	0.857**	0.269**	0.186†
1	3	0.289**	0.095**	0.076†
1	4	1.315**	0.497**	0.314†

(b) CNN

Table Numb.	Model Numb.	Vocal Pitch Coeff.	Overall Sent. Coeff.	Cong. Sent. Coeff.
1	1	0.587**	0.030	0.055
1	2	1.284**	0.129	0.709*
1	3	–	–	–
1	4	0.563**	0.073	0.080

(c) Fox News

Table Numb.	Model Numb.	Vocal Pitch Coeff.	Overall Sent. Coeff.	Cong. Sent. Coeff.
1	1	0.571**	0.198*	0.173
1	2	0.855**	0.212†	0.108
1	3	0.229**	0.068†	0.046
1	4	1.558**	0.485*	0.375

(d) MSNBC

Table Numb.	Model Numb.	Vocal Pitch Coeff.	Overall Sent. Coeff.	Cong. Sent. Coeff.
1	1	0.537**	0.369**	0.297*
1	2	0.792**	0.313**	0.145
1	3	0.228**	0.131**	0.115*
1	4	0.503**	0.376**	0.241

Note: In the first two columns we report the corresponding table and model number from the main text. In columns 3-6 we report the coefficient and level of statistical significance for each of our main independent variables. In Panel A, we report these estimates pooling all network data together. In Panels B, C, and D, we restrict our data to only include CNN, Fox News, and MSNBC broadcasts respectively. Levels of significance (with standard errors clustered around each MC) are reported as follows: †p < .1; *p < .05; **p < .01.

predictors regardless of the dependent variable suggesting these results cannot be attributed to a single network.

Overall, `Vocal Pitch` is a positive and statistically significant predictor in every network model we estimated. This provides strong support for our first hypothesis. `Overall Sentiment` is also statistically significant in 7 models which yields partial support for our second hypotheses. Finally, we find the least amount of support for our third hypothesis. Here, we find `Congress Sentiment` is only statistically significant in 3 of 11 models which is probably why we find the least amount of statistical support for this hypothesis in the main text. With that said, most floor speeches do not appear on television which means disaggregating the data can significantly decrease our degrees of freedom. This is one of the reasons why we report pooled models in the main text.

S7.2 Each Independent Variable By Itself

Along these lines, in Table S11 we report models in which `Vocal Pitch`, `Overall Sentiment`, and `Congress Sentiment` are included in isolation. For example, in Panel A we re-estimated all of the models in Table 1 *including* `Vocal Pitch` and *excluding* `Overall Sentiment` and `Congress Sentiment`. In the first two columns of each panel, we report the corresponding table and model number from the main text, meaning the first row corresponds with Table 1, Model 1 except `Overall Sentiment` and `Congress Sentiment` are no longer included as independent variables. In the third and fourth columns, we report the coefficient, standard error, and level of significance for the variable of interest. In Panel A, this is `Vocal Pitch` which means the last two columns only report the coefficient for `Vocal Pitch`. Finally, the models highlighted in grey include the control variables we reported in the main text whereas in the unhighlighted rows we do not include any controls.

Table S11 shows our results actually improve when our main independent variables are included by themselves in the models reported in Table 1. For example, `Congress Sentiment` is now consistently statistically significant at the 0.05-level whereas in the main text this same variable is only statistically significant at the 0.10-level. To check whether this is due to our main independent variables being correlated with one another, we report the correlation matrix for `Vocal Pitch`, `Overall Sentiment`, and `Congress Sentiment` in Table S12. Here, we report the correlation between the variables reported in the associated row and column. Grey cells indicated the correlation is statistically significant at the 0.05-level. Even though `Vocal Pitch` is weakly correlated with `Overall Sentiment`, the rest of the correlations approximate zero which is why we included all four variables in the models we reported in Table 1.

S7.3 Zero-Inflated Models

Given that very few floor speeches are actually aired, Models 1.1 and 1.2 in the main text were estimated using a rare logistic regression. For Models 1.3 and 1.4, we report estimates

Table S11: CNN, Fox News, and MSNBC Coverage of Verbal and Non-Verbal Emotional Expressions on the Floor of the U.S. House of Representatives (Each Independent Variable By Itself)

(a) Vocal Pitch

Table Number	Model Number	Coefficient	Standard Error
1	1	0.643	(0.055)**
1	2	0.632	(0.055)**
1	3	0.877	(0.087)**
1	4	0.909	(0.092)**
1	5	0.300	(0.041)**
1	6	0.292	(0.038)**
1	7	1.369	(0.197)**
1	8	1.329	(0.199)**

(b) Overall Sentiment

Table Number	Model Number	Coefficient	Standard Error
1	1	0.345	(0.053)**
1	2	0.332	(0.052)**
1	3	0.397	(0.083)**
1	4	0.449	(0.098)**
1	5	0.147	(0.031)**
1	6	0.146	(0.031)**
1	7	0.729	(0.139)**
1	8	0.730	(0.139)**

(c) Congress Sentiment

Table Number	Model Number	Coefficient	Standard Error
1	1	0.237	(0.103)*
1	2	0.237	(0.102)*
1	3	0.203	(0.087)*
1	4	0.176	(0.100)†
1	5	0.096	(0.041)*
1	6	0.095	(0.043)*
1	7	0.412	(0.169)*
1	8	0.407	(0.178)*

Note: In the first two columns we report the corresponding table and model number from the main text. In the third and fourth columns, we report the coefficient and standard error when the an independent variable is estimated in isolation. For example, the first row in Panel A reports the coefficient and standard error from Table 1, Model 1 when `Vocal Pitch` is *included* and `Overall/Congress Sentiment` are excluded as predictors. Levels of significance (with standard errors clustered around each MC) are reported as follows: $^{\dagger}p < .1$; $^{*}p < .05$; $^{**}p < .01$.

Table S12: Correlation Between Verbal and Non-Verbal Emotional Expressions

	Vocal Pitch	Overall Sentiment	Congress Sentiment
Vocal Pitch	–	0.19	0.02
Overall Sentiment	0.19	–	0.08
Congress Sentiment	0.02	0.08	–

Note: Correlation matrix for all three independent variables reported in Table 1. Pearson's ρ is statistically significant at the 0.05-level in the grey cells, otherwise no significant correlation is found.

from negative binomial regressions. Unfortunately, Zelig does not support zero-inflated count models which is why they are not included in Table 1 in the main text. To determine whether this would have changed our substantive results, we re-estimated Models 1.3 and 1.4 using zero-inflated negative binomial regressions.

In Tables S13 and S14, we report the results from our zero-inflated models. In Models 1 and 2, we provide the corresponding results from Table 1 in the main text. Since one can include any number of parameters in the logistic portion of the model, we estimated a version of the zero-inflated models where the intercept was included as the only variable in that portion of the model. These results are shown in Models 3 and 4. In the last two models (5 and 6), we include all variables in both the count and logistic portions of the zero-inflated models.

Beginning with the count portion of the model (see Table S13), we find many of the results are the same as those reported in the main text. For example, `Vocal Pitch` is a statistically significant ($p < 0.01$) predictor in three of the four zero-inflated models. We also find that when the logistic portion of the model only includes an intercept, `Overall Sentiment` is also statistically significant at the 0.01-level. Similarly, just like in the main text, we find the least statistical support for our third hypothesis since `Congress Sentiment` is never statistically significant at the 0.05-level. This gives us strong evidence in support of our first hypothesis and some partial support for our second hypothesis.

We find similar results in the logistic component of the zero-inflated models. Here, it is important to remember that the logistic portion of the model predicts non-occurrence, meaning positive coefficients imply the floor speech is *less* likely to appear on television. Given that, the negative and statistically significant coefficient associated with *Vocal Pitch* provides additional support for our first hypothesis since it suggests increased vocal pitch decreases the likelihood of non-occurrence. Said differently, a floor speech is *more* likely to appear on television when it is delivered at a higher vocal pitch.

With this in mind, the rest of the results are similar to those reported in the main text. Although less pronounced than `Vocal Pitch`, the coefficient associated with *Overall Sentiment* is statistically significant at the 0.05-level when other variables are included as controls. Similarly, *Congress Sentiment* never reaches traditional levels of statistical significance which again suggests our third hypothesis has the least amount of empirical support. With that said, it is important to note that the logistic component of the zero-inflated negative binomial regression does not use a penalized likelihood function similar to `ReLogit` which is why we did not report these models in the main text.

Ultimately, we encourage scholars to report the results from a variety of different modeling approaches. As we explain in footnote 7 in the main text, we report the results from 89 different models. In 87 of those models (or 97.75 percent) `Vocal Pitch` was a statistically significant ($p < 0.05$) predictor of television coverage. In 76 of those models (or 85.39 percent) `Overall Sentiment` was statistically significant at the 0.05-level which is why we do not think our main results can be attributed to any single modeling choice. Indeed, regardless of the approach we use, `Vocal Pitch` and `Overall Sentiment` are consistently

Table S13: Re-Estimating the Models in Table 1 Using a Zero-Inflated Negative Binomial Regression (Count Component)

	Dependent variable:					
	Total Coverage					
				Zero-Inflated		
	Negative Binomial			*Negative Binomial*		
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	-5.421^{**}	-6.293^{**}	-5.418^{**}	-6.293^{**}	-3.241	-3.367^{**}
	(0.139)	(0.433)	(0.249)	(0.483)	(2.051)	(1.206)
Vocal Pitch	0.857^{**}	0.882^{**}	0.857^{**}	0.882^{**}	0.178	0.631^{**}
	(0.093)	(0.101)	(0.093)	(0.104)	(0.296)	(0.210)
Overall Sentiment	0.269^{**}	0.318^{**}	0.269^{**}	0.318^{**}	-0.061	-0.083
	(0.090)	(0.106)	(0.097)	(0.113)	(0.214)	(0.142)
Congress Sentiment	$0.186^{\dagger}$	0.142	$0.186^{\dagger}$	0.142	0.077	-0.047
	(0.106)	(0.115)	(0.096)	(0.109)	(0.105)	(0.109)
Republican		$-0.677^{\dagger}$		$-0.677^{\dagger}$		-0.638
		(0.358)		(0.363)		(0.542)
|DW-Nominate|		1.745^{**}		1.745^{*}		2.135
		(0.604)		(0.681)		(1.303)
Party Issue		-0.134		-0.134		0.195
		(0.101)		(0.097)		(0.154)
Seniority		$0.016^{\dagger}$		0.016		-0.040
		(0.008)		(0.010)		(0.025)
House Leader		2.090^{**}		2.090^{**}		1.000^{**}
		(0.425)		(0.427)		(0.327)
Committee Chair		0.252		0.252		0.267
		(0.304)		(0.321)		(0.477)
Male		0.204		0.204		0.542
		(0.297)		(0.253)		(0.422)
White		0.173		0.173		-0.374
		(0.382)		(0.371)		(0.465)
CQ Bills		$0.195^{\dagger}$		0.195		-1.621^{**}
		(0.100)		(0.146)		(0.377)
One Minute		0.070		0.070		0.036
		(0.292)		(0.269)		(0.393)
Duration		-0.160^{*}		-0.160^{*}		0.077
		(0.080)		(0.068)		(0.087)
Election Year		-1.015^{**}		-1.015^{**}		0.426
		(0.330)		(0.272)		(0.488)
N	71,198	71,197	71,198	71,197	71,198	71,197
Log-Lik	$-1{,}965.667$	$-1{,}909.634$	$-1{,}964.668$	$-1{,}908.634$	$-1{,}941.947$	$-1{,}824.213$
AIC	3,939.335	3,851.267	3,941.336	3,853.267	3,901.893	3,714.426

Note: The dependent variable is reported above the corresponding columns. All coefficients are estimated using a zero-inflated negative binomial regression. Please refer to pages S49–S52 in the Supplemental Information for more details. `Vocal Pitch` and `Overall/Congress Sentiment` are scaled to standard deviations above or below the MC's baseline. Levels of significance are reported as follows: $^{\dagger}p < .1$; $^{*}p < .05$; $^{**}p < .01$. Clustered (around each MC) standard errors are reported in parentheses.

Table S14: Re-Estimating the Models in Table 1 Using a Zero-Inflated Negative Binomial Regression (Zero Component)

	Dependent variable:					
	Total Coverage					
	Rare Events		*Zero-Inflated*			
	Logistic Regression		*Negative Binomial*			
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	-5.883^{**}	-6.013^{**}	-5.922	-14.142	2.000	2.723^{*}
	(0.074)	(0.307)	(83.905)	(43.879)	(2.373)	(1.225)
Vocal Pitch	0.617^{**}	0.605^{**}			-0.887^{**}	$-0.395^{\dagger}$
	(0.057)	(0.058)			(0.270)	(0.215)
Overall Sentiment	0.237^{**}	0.226^{**}			-0.333	-0.404^{*}
	(0.061)	(0.058)			(0.224)	(0.163)
Congress Sentiment	$0.181^{\dagger}$	$0.177^{\dagger}$			-0.172	$-0.320^{\dagger}$
	(0.102)	(0.102)			(0.158)	(0.166)
Republican		$-0.332^{\dagger}$				-0.015
		(0.200)				(0.597)
|DW-Nominate|		1.197^{**}				0.078
		(0.415)				(1.330)
Party Issue		-0.073				0.339
		(0.063)				(0.215)
Seniority		0.022^{**}				-0.067^{**}
		(0.006)				(0.022)
House Leader		1.899^{**}				-2.040^{*}
		(0.195)				(0.829)
Committee Chair		0.197				-0.042
		(0.223)				(0.553)
Male		-0.222				0.661
		(0.181)				(0.414)
White		-0.195				-0.276
		(0.230)				(0.462)
CQ Bills		0.212^{**}				-3.013^{**}
		(0.051)				(0.800)
One Minute		-0.053				-0.076
		(0.172)				(0.442)
Duration		-0.116^{*}				$0.224^{\dagger}$
		(0.048)				(0.120)
Election Year		-1.005^{**}				1.975^{**}
		(0.174)				(0.440)
N	71,198	71,197	74,151	74,151	71,198	71,197
Log-Lik	$-1{,}624.930$	$-1{,}541.353$	$-1{,}718.063$	$-1{,}718.063$	$-1{,}624.930$	$-1{,}541.353$
AIC	3,257.860	3,114.706	3,941.336	3,853.267	3,901.893	3,714.426

Note: The dependent variable is reported above the corresponding columns. All coefficients are estimated using a zero-inflated negative binomial regression. Please refer to pages S49–S52 in the Supplemental Information for more details. `Vocal Pitch` and `Overall/Congress Sentiment` are scaled to standard deviations above or below the MC's baseline. Levels of significance are reported as follows: $^{\dagger}p < .1$; $^{*}p < .05$; $^{**}p < .01$. Clustered (around each MC) standard errors are reported in parentheses.

found to be significant predictors of the quality and quantity of television coverage.

S7.4 Excluding Potential Outliers

To address concerns that our results are being driven by a handful of extreme speeches, we also re-estimated all our models excluding potential outliers. To do so, we use a conservative definition of what constitutes an outlier: any speech with a vocal pitch more than ± 2 standard deviations away from a speaker's baseline. We did the same for the number of televisions appearances. Here, we define any speech that was aired more than 10 times as an "outlier." Ultimately, we find the results we report in the main text remain unchanged even after restricting our data in this way.

In Table S15, we report the results excluding either vocal pitch or television coverage outliers. In the first two columns, we report the corresponding table and model number from the main text. The third column reports the results when extreme values of vocal pitch are excluded from the data. The fourth column does the same for television coverage. Each panel shows the independent variable of interest, meaning in the first row of Panel A we show the coefficient and level of significance for `Vocal Pitch` when either vocal pitch or television coverage outliers are excluded across all the models reported in Table 1 of the main text.

Regardless of the independent variable, excluding extreme values does not affect our substantive results. Indeed, `Vocal Pitch`, `Overall Sentiment`, and `Congress Sentiment` are all in the same direction and have the same level of significance as the coefficients reported in Table 1. In fact the only time the coefficients drop below traditional levels of statistical significance is when Models 1.3 and 1.4 are re-estimated excluding speeches in which the speaker spoke greater than ± 2 standard deviations away from his/her baseline. Here, the coefficient for `Congress Sentiment` is no longer statistically significant at the 0.10-level (see Panel C, Models 3-4) even though the direction of the relationship is the same. Despite this result, we think Table S15 provides strong evidence that our substantive results cannot be attributed to a handful of influential observations.

S7.5 Imputing vs. Not Imputing Missing Observations

As we explained in footnote 4 in the main text, for the congressional sentiment scores we assigned a zero to any speech in which Congress was not referenced. Since `Congress Sentiment` is standardized using each MC's baseline, this equates to saying the MC was neither more or less negative towards Congress in those speeches. We think this imputation makes sense, but we understand that the missing values were not randomly assigned so any imputation makes strong assumptions about the data generating process. Given that, we re-estimated all of our models excluding cases in which Congress was not mentioned. These results are shown in Table S16.

In the first two columns of each panel in Table S16, we report the corresponding table

Table S15: CNN, Fox News, and MSNBC Coverage of Verbal and Non-Verbal Emotional Expressions on the Floor of the U.S. House of Representatives (Excluding Outliers)

(a) Vocal Pitch

Table Number	Model Number	Coefficient Excluding Vocal Pitch Outliers	Coefficient Excluding TV Coverage Outliers
1	1	0.914**	0.613**
1	2	0.896**	0.601**
1	3	0.975**	0.810**
1	4	1.050**	0.815**
1	5	0.390**	0.290**
1	6	0.388**	0.282**
1	7	1.693**	1.109**
1	8	1.696**	1.056**

(b) Overall Sentiment

Table Number	Model Number	Coefficient Excluding Vocal Pitch Outliers	Coefficient Excluding TV Coverage Outliers
1	1	0.164*	0.241**
1	2	0.157*	0.229**
1	3	0.257**	0.232**
1	4	0.317**	0.254**
1	5	0.072*	0.097**
1	6	0.074*	0.097**
1	7	0.391**	0.423**
1	8	0.407**	0.422**

(c) Congress Sentiment

Table Number	Model Number	Coefficient Excluding Vocal Pitch Outliers	Coefficient Excluding TV Coverage Outliers
1	1	0.225*	0.184†
1	2	0.230*	0.180†
1	3	0.203	0.207†
1	4	0.144	0.190†
1	5	0.094*	0.078†
1	6	0.092*	0.077†
1	7	0.413*	0.279†
1	8	0.410*	0.279†

Note: In the first two columns we report the corresponding table and model number from the main text. In the third column, we report the coefficient and level of statistical significance excluding potential vocal pitch outliers. To do so, we use a conservative definition of what constitutes an outlier: any speech with a vocal pitch more than ± 2 standard deviations away from a speaker's baseline. In the fourth column, we do the same for television coverage. Here, we define any speech that was aired more than 10 times as an "outlier." Each panel shows the independent variable of interest, meaning in the first row of Panel A we show the coefficient and level of significance for `Vocal Pitch` when outliers are excluded for either vocal pitch or television coverage. Levels of significance (with standard errors clustered around each MC) are reported as follows: $^{\dagger}p < .1$; $^{*}p < .05$; $^{**}p < .01$.

Table S16: CNN, Fox News, and MSNBC Coverage of Verbal and Non-Verbal Emotional Expressions on the Floor of the U.S. House of Representatives (Imputing Missing Observations)

(a) Vocal Pitch

Table Number	Model Number	Coefficient `NA`s Imputed	Coefficient `NA`s Excluded
1	1	0.617**	0.479**
1	2	0.605**	0.459**
1	3	0.857**	0.564**
1	4	0.882**	0.616**
1	5	0.289**	0.272**
1	6	0.281**	0.263**
1	7	1.315**	1.065**
1	8	1.272**	1.007**

(b) Overall Sentiment

Table Number	Model Number	Coefficient `NA`s Imputed	Coefficient `NA`s Excluded
1	1	0.237**	0.288*
1	2	0.226**	0.259*
1	3	0.269**	0.346†
1	4	0.318**	0.460*
1	5	0.095**	0.130
1	6	0.095**	0.131†
1	7	0.497**	0.747*
1	8	0.502**	0.732**

(c) Congress Sentiment

Table Number	Model Number	Coefficient `NA`s Imputed	Coefficient `NA`s Excluded
1	1	0.181†	0.142
1	2	0.177†	0.139
1	3	0.186†	0.142†
1	4	0.142	0.099
1	5	0.076†	0.077
1	6	0.075†	0.080
1	7	0.314†	0.259
1	8	0.310†	0.254

Note: In the first two columns we report the corresponding table and model number from the main text. In the third column, we report the coefficient and level of statistical significance replacing all `NA`s in `Congress Sentiment` with 0s. These are the coefficients reported in Table 1. In the fourth column, we report the coefficient and level of statistical significance excluding the `NA`s in `Congress Sentiment`. Since `Vocal Pitch`, `Overall Sentiment`, and `Congress Sentiment` are always included in the same model, this equates to only considering speeches in which Congress was mentioned. Each panel shows the independent variable of interest, meaning in the first row of Panel A we show the coefficient and level of significance for `Vocal Pitch` when the `NA`s in `Congress Sentiment` are either excluded or imputed. Levels of significance (with standard errors clustered around each MC) are reported as follows: $^{\dagger}p < .1$; $^{*}p < .05$; $^{**}p < .01$.

and model number from the main text. The third column reports the results when zeros are used to replace the NAs in Congress Sentiment when Congress is not mentioned in a speech. These coefficients are identical to those reported in Table 1 in the main text. In the last column, we report coefficients from the same models when the NAs in Congress Sentiment are excluded. Since Vocal Pitch, Overall Sentiment, and Congress Sentiment are included in all models, this equates to only considering speeches in which Congress was mentioned.

In total, 25,631 speeches mentioned Congress and 48,520 speeches did not mention Congress. When the former are the only speeches used in the analysis, the results for Vocal Pitch and Overall Sentiment largely remain the same, but the coefficient for Congress Sentiment is no longer statistically significant at the 0.10-level. Although the interpretation of these results is the same as those in the main text, we decided to report the imputed results because we were uncomfortable eliminating 65.43 percent of our data and only using speeches in which Congress was mentioned.

In earlier drafts of this manuscript, we presented separate models for each variable by itself (which are now reported in Table S11), but we were told it would be better to include all three variables in the same model. However, this made it more difficult to deal with the NAs in Congress Sentiment which was not a concern for either Vocal Pitch or Overall Sentiment since neither of those variables had missing values. Unfortunately, we could not find a straightforward way to account for the missing values in Congress Sentiment while simultaneously using all of our data to estimate the effect of the other independent variables.

For us, zero imputation makes sense, but there should be a broader discussion about what scholars should do in these instances. However, this is beyond the scope of the present paper. Regardless, scholars should report the results with and without imputation which is why we include Table S16 in the Supplemental Information and reference the results in footnote 4 in the main text.

S7.6 R vs. Stata

On page 18 of the main text, we say "All models were estimated using Zelig with standard errors clustered around each MC to account for speech-level dependence." As we explain in footnote 7, we report clustered standard errors primarily because there currently is no multilevel implementation of ReLogit. To add additional robustness to our results, in Table S17 we re-estimated all of the models reported in Table 1 using the glmer, glmer.nb, and lmer functions from the lme4 library in the R statistical language.

As you can see, Vocal Pitch is a statistically significant predictor in every model, even though Models 2 and 4 did not converge when a randomly varying intercept was included for each MC. This provides additional support for our first hypothesis, although the same cannot be said for our second hypothesis. Here, Overall Sentiment is statistically significant at the 0.05-level in Models 1-4, but not statistically significant in Models 5-8. Given the unreasonably large log-likelihoods, we suspect this may have to do with the model estimation

Table S17: Re-Estimating the Models in Table 1 Including a Randomly Varying Intercept for Each MC (Estimated in `R`)

	Televised		Total Coverage		Total Minutes		Total Viewers	
	`glmer`		`glmer.nb`		`lmer`		`lmer`	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	-6.622^{**}	-6.978^{**}	-6.717^{**}	-7.658^{**}	0.001^{**}	0.001	0.005^{**}	0.002
	(0.153)	(0.534)	(0.212)	(0.001)	(0.0002)	(0.001)	(0.001)	(0.003)
Vocal Pitch	0.644^{**}	0.631^{**}	0.808^{**}	0.824^{**}	0.001^{**}	0.001^{**}	0.005^{**}	0.005^{**}
	(0.053)	(0.053)	(0.077)	(0.001)	(0.0002)	(0.0002)	(0.001)	(0.001)
Overall Sentiment	0.242^{**}	0.232^{**}	0.264^{**}	0.283^{**}	0.0001	0.0001	0.001	0.001
	(0.067)	(0.064)	(0.084)	(0.001)	(0.0002)	(0.0002)	(0.001)	(0.001)
Congress Sentiment	$0.154^{\dagger}$	$0.174^{\dagger}$	0.169	0.195^{**}	0.0004	0.0004	0.001	0.001
	(0.087)	(0.089)	(0.110)	(0.001)	(0.0003)	(0.0003)	(0.001)	(0.001)
Republican		−0.458		-0.687^{**}		−0.001		-0.004^{*}
		(0.316)		(0.001)		(0.001)		(0.002)
|DW-Nominate|		1.539^{*}		2.429^{**}		0.0005		$0.008^{\dagger}$
		(0.679)		(0.001)		(0.001)		(0.004)
Party Issue		−0.074		-0.089^{**}		−0.0001		−0.0001
		(0.065)		(0.001)		(0.0002)		(0.001)
Seniority		0.031^{**}		0.026^{**}		0.0001^{*}		0.00003
		(0.010)		(0.001)		(0.00002)		(0.0001)
House Leader		2.628^{**}		3.267^{**}		0.016^{**}		0.034^{**}
		(0.568)		(0.001)		(0.002)		(0.005)
Committee Chair		0.141		0.371^{**}		0.001		0.001
		(0.277)		(0.001)		(0.001)		(0.002)
Male		0.026		0.158^{**}		−0.0004		0.003
		(0.269)		(0.001)		(0.001)		(0.002)
White		−0.128		-0.043^{**}		−0.0003		−0.0004
		(0.347)		(0.001)		(0.001)		(0.003)
CQ Bills		0.191^{*}		0.190^{**}		0.001^{**}		−0.0004
		(0.082)		(0.001)		(0.0003)		(0.001)
One Minute		−0.081		0.026^{**}		−0.00004		−0.001
		(0.167)		(0.001)		(0.0004)		(0.001)
Duration		-0.155^{**}		-0.159^{**}		$0.0001^{\dagger}$		−0.0003
		(0.045)		(0.001)		(0.00004)		(0.0002)
Election Year		-1.017^{**}		-1.336^{**}		−0.0005		-0.004^{**}
		(0.175)		(0.001)		(0.0003)		(0.001)
N	71,198	71,197	71,198	71,197	71,198	71,197	71,198	71,197
Log Lik	−1,559.009	−1,506.536	−1,908.603	−1,857.343	126,181.800	126,144.100	27,999.270	27,960.740
AIC	3,128.018	3,047.071	3,829.206	3,750.686	−252,351.600	−252,252.300	−55,986.530	−55,885.480

Note: All models estimated with a randomly varying intercept for each MC. The dependent variable is listed above each column. Immediately below we list the `R` function we used. All models converged except for Models 2 and 4 which **did not converge**. These models are highlighted in grey. `Vocal Pitch` and `Overall/Congress Sentiment` are scaled to standard deviations above or below the MC's baseline. Levels of significance are reported as follows: $^{\dagger}p < .1$; $^{*}p < .05$; $^{**}p < .01$.

in R, especially since lme4 does not have a Tobit implementation which is what we used in Table 1. For these reasons, we re-estimated all of the multilevel models in Table S17 using Stata. These are reported in Table S18.

Similar to Table S17, in each of these models we included a randomly varying intercept for each MC. Ultimately, we found the Stata models performed a lot better. Not only were there no convergence issues, but all of the log likelihoods seem reasonable. Moreover, the results are essentially the same as those that are reported in Table 1 with both Vocal Pitch and Overall Sentiment always yielding statistically significant results. Unlike R, Stata also has a multilevel Tobit implementation which is reported in Models 5-8 leading us to believe Stata is perhaps better suited at estimating the multilevel versions of the models we reported in the main text.

Given the noticeable differences between Tables S17 and S18, we decided to investigate why Stata and R performed differently which led us to a post on Andrew Gelman's blog titled "R vs. Stata, or, Different ways to estimate multilevel models." In that post,[S6] a reader named Cyrus essentially asked why Stata and R yielded different multilevel results. More specifically, both the reader and Andrew Gelman noted that glmer's default in R for the psuedo-likelihood estimation was an "adaptive Gaussian quadrature with 1 integration point" whereas Stata uses "a very sensible 7-point quadrature default." Although Gelman does not choose a side, he does note the author of the Stata function says it is superior to glmer and does not seem to quibble too much with this assertion.

For us, these results suggests scholars should present multilevel results using both Stata and R given the different estimation approaches. Given the multilevel differences we found, we also re-estimated all of the models reported in Table 1 in Stata using robust clustered standard errors. These results are reported in Table Table S19. Ultimately, we found essentially the same results even though we were using an interval regression instead of a Tobit regression in Models 5-8. Similarly, in Table S20 we report the results from a Firth logistic regression instead of ReLogit. In both instances, Stata either did not allow for robust clustered standard errors or did not include the exact same model as R. However, the results in Tables S19 and S20 are so similar to Table 1 we think it is safe to say our choice of statistical software also cannot fully account for our main results.

References

Bachorowski, Jo-Anne and Michael J. Owren. 1995. "Vocal Expression of Emotion: Acoustic Properties of Speech are Associated with Emotional Intensity and Context." *Psychological Science* 6(4):219–224. S19

Bänziger, Tanja and Klaus R Scherer. 2007. Using actor portrayals to systematically study multimodal emotion expression: The GEMEP corpus. In *International conference on affective computing and intelligent interaction*. Springer pp. 476–487. S21

[S6] https://andrewgelman.com/2010/09/10/r_vs_stata_or_d/

Table S18: Re-Estimating the Models in Table 1 Including a Randomly Varying Intercept for Each MC (Estimated in `Stata`)

	Televised		Total Coverage		Total Minutes		Total Viewers	
	`xtmelogit`		`menbreg`		`metobit`		`metobit`	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	-6.525^{**}	-6.931^{**}	-6.397^{**}	-7.315^{**}	-3.494^{**}	-3.731^{**}	-16.240^{**}	-16.629^{**}
	(0.140)	(0.511)	(0.179)	(0.674)	(0.197)	(0.288)	(0.986)	(1.374)
Vocal Pitch	0.643^{**}	0.634^{**}	0.870^{**}	0.885^{**}	0.284^{**}	0.283^{**}	1.317^{**}	1.294^{**}
	(0.053)	(0.053)	(0.083)	(0.085)	(0.028)	(0.029)	(0.136)	(0.135)
Overall Sentiment	0.242^{**}	0.232^{**}	0.284^{**}	0.304^{**}	0.096^{**}	0.097^{**}	0.507^{**}	0.510^{**}
	(0.067)	(0.064)	(0.090)	(0.090)	(0.029)	(0.029)	(0.141)	(0.138)
Congress Sentiment	$0.155^{\dagger}$	$0.171^{\dagger}$	0.180	$0.207^{\dagger}$	$0.067^{\dagger}$	$0.076^{\dagger}$	0.291	0.325^{**}
	(0.087)	(0.090)	(0.119)	(0.122)	(0.038)	(0.039)	(0.183)	(0.188)
Republican		$-0.504^{\dagger}$		$-0.670^{\dagger}$		-0.201		-1.082^{*}
		(0.302)		(0.393)		(0.123)		(0.586)
DW-Nominate		1.626^{*}		2.374^{**}		0.627^{*}		3.005^{*}
		(0.652)		(0.850)		(0.268)		(1.274)
Party Issue		-0.076		-0.096		-0.033		-0.146
		(0.065)		(0.084)		(0.028)		(0.134)
Seniority		0.031^{**}		$0.024^{\dagger}$		0.012^{**}		0.050^{**}
		(0.010)		(0.013)		(0.004)		(0.019)
House Leader		2.604^{**}		3.156^{**}		1.126^{**}		5.241^{**}
		(0.524)		(0.799)		(0.228)		(1.069)
Committee Chair		0.132		0.378		0.079		0.397
		(0.273)		(0.353)		(0.113)		(0.552)
Male		-0.001		0.188		0.007		0.202
		(0.254)		(0.333)		(0.104)		(0.492)
White		-0.096		-0.062		-0.027		-0.239
		(0.330)		(0.440)		(0.136)		(0.631)
CQ Bills		0.196^{*}		0.208		0.096^{*}		0.286
		(0.081)		(0.139)		(0.037)		(0.189)
One Minute		-0.071		0.043		0.029		-0.224
		(0.167)		(0.210)		(0.070)		(0.340)
Duration		-0.152^{**}		-0.164^{**}		-0.037^{*}		-0.320^{**}
		(0.045)		(0.051)		(0.016)		(0.093)
Election Year		-1.023^{**}		-1.411^{**}		-0.405^{**}		-2.168^{**}
		(0.175)		(0.215)		(0.074)		(0.372)
N	71,198	71,197	71,198	71,197	71,198	71,197	71,198	71,197
Log Lik	−1,564.994	−1,511.022	−1,913.696	−1,863.206	−1,565.799	−1,516.125	−1,780.505	−1,726.19
AIC	3,139.989	3,056.043	3,839.392	3,762.412	3,143.598	3,068.249	3,573.010	3,488.380

Note: All models estimated with a randomly varying intercept for each MC. The dependent variable is listed above each column. Immediately below we list the `Stata` function we used. `Vocal Pitch` and `Overall/Congress Sentiment` are scaled to standard deviations above or below the MC's baseline. Levels of significance are reported as follows: $^{\dagger}p < .1$; $^{*}p < .05$; $^{**}p < .01$.

Table S19: Re-Estimating the Models in Table 1 Using Robust Clustered Standard Errors in `Stata`

	Televised		Total Coverage		Total Minutes		Total Viewers	
	`logit`		`nbreg`		`intreg`		`intreg`	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	-5.889^{***}	-6.040^{***}	-5.421^{***}	-6.293^{***}	-3.512^{***}	-3.548^{***}	-16.053^{***}	-15.724^{***}
	(0.113)	(0.310)	(0.140)	(0.485)	(0.442)	(0.436)	(1.834)	(2.073)
Vocal Pitch	10.217^{***}	10.039^{***}	4.975^{***}	5.117^{***}	0.289^{***}	0.281^{***}	1.315^{***}	1.272^{***}
	(0.059)	(0.058)	(0.094)	(0.104)	(0.040)	(0.037)	(0.194)	(0.196)
Overall Sentiment	3.915^{***}	3.733^{***}	1.556^{***}	1.839^{***}	0.095^{***}	0.095^{***}	0.497^{***}	0.502^{***}
	(0.070)	(0.066)	(0.099)	(0.117)	(0.033)	(0.033)	(0.145)	(0.141)
Congress Sentiment	1.663^{**}	1.626^{*}	0.627^{*}	0.478	0.076^{*}	0.075^{*}	0.314^{*}	0.310^{*}
	(0.085)	(0.086)	(0.096)	(0.110)	(0.042)	(0.045)	(0.170)	(0.181)
Republican		−2.784		-1.971^{*}		−0.132		−0.853
		(0.272)		(0.363)		(0.112)		(0.568)
|DW - Nominate|		4.391^{**}		2.232^{**}		0.443^{**}		2.242^{**}
		(0.499)		(0.681)		(0.206)		(1.023)
Party Issue		−1.225		−0.777		−0.030		−0.130
		(0.062)		(0.098)		(0.028)		(0.130)
Seniority		3.574^{***}		0.908		0.010^{***}		0.034^{**}
		(0.007)		(0.010)		(0.004)		(0.016)
House Leader		5.337^{***}		2.057^{***}		0.857^{***}		3.964^{***}
		(0.314)		(0.427)		(0.202)		(0.807)
Committee Chair		0.826		0.403		0.099		0.352
		(0.266)		(0.318)		(0.117)		(0.562)
Male		−1.395		0.458		−0.075		−0.101
		(0.285)		(0.259)		(0.121)		(0.493)
White		−0.857		0.286		−0.065		−0.366
		(0.277)		(0.371)		(0.130)		(0.548)
CQ Bills		1.997^{***}		0.667		0.109^{***}		0.329^{**}
		(0.057)		(0.145)		(0.031)		(0.138)
One Minute		−0.483		0.180		0.025		−0.254
		(0.146)		(0.269)		(0.063)		(0.305)
Duration		-7.623^{**}		-3.529^{**}		−0.003		-0.274^{**}
		(0.051)		(0.068)		(0.019)		(0.114)
Election Year		-8.025^{***}		-2.808^{***}		-0.402^{***}		-2.122^{***}
		(0.181)		(0.261)		(0.086)		(0.373)
N	71,198	71,197	71,198	71,197	71,198	71,197	71,198	71,197
Log Lik	−1,624.930	−1,541.353	−1,964.667	−1,908.634	−1,627.715	−1,545.823	−1,830.763	−1,750.996
AIC	3,257.860	3,114.706	3,939.335	3,851.267	3,265.430	3,125.646	3,671.525	3,535.993

Note: The dependent variable is listed above each column. Immediately below we list the `Stata` function we used. `Vocal Pitch` and `Overall/Congress Sentiment` are scaled to standard deviations above or below the MC's baseline. Levels of significance are reported as follows: $^{\dagger}p < .1$; $^{*}p < .05$; $^{**}p < .01$. Clustered (around each MC) robust standard errors reported in parentheses. These were estimated using the `robust` and `cluster` options in `Stata`.

Table S20: Re-Estimating the Models in Table 1 Using Robust Clustered Standard Errors in `Stata`

	Dependent variable:	
	Televised	
	`firthlogit`	
	(1)	(2)
Constant	-5.883^{**}	-6.014^{**}
	(0.075)	(0.354)
Vocal Pitch	0.617^{**}	0.605^{**}
	(0.051)	(0.051)
Overall Sentiment	0.237^{**}	0.226^{**}
	(0.066)	(0.064)
Congress Sentiment	0.181^{*}	0.177^{*}
	(0.087)	(0.088)
Republican		-0.332
		(0.214)
|DW-Nominate|		1.197^{*}
		(0.473)
Party Issue		-0.073
		(0.065)
Seniority		0.022^{**}
		(0.007)
House Leader		1.898^{**}
		(0.191)
Committee Chair		0.197
		(0.231)
Male		-0.222
		(0.169)
White		-0.194
		(0.226)
CQ Bills		0.212^{**}
		(0.079)
One Minute		-0.053
		(0.163)
Duration		-0.116^{**}
		(0.043)
Election Year		-1.005^{**}
		(0.172)
Observations	71,198	71,197
Log Lik	$-1{,}613.997$	$-1{,}503.774$
AIC	3,235.993	3,039.548

Note: The dependent variable is listed above each column. Immediately below we list the `Stata` function we used. `Vocal Pitch` and `Overall/Congress Sentiment` are scaled to standard deviations above or below the MC's baseline. Levels of significance are reported as follows: $^{\dagger}p < .1$; $^{*}p < .05$; $^{**}p < .01$. Clustered (around each MC) robust standard errors reported in parentheses. These were estimated using the `robust` and `cluster` options in `Stata`.

Blei, David M., Andrew Y. Ng and Michael I. Jordan. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3:993–1022. S42

Boersma, Paul. 1993. "Accurate Short-Term Analysis of the Fundamental Frequency and the Harmonics-to-Noise Ratio of a Sampled Sound." *Proceedings of the Institute of Phonetic Sciences* 17:97–110. S5

DePaulo, Bella M., James J. Lindsay, Brian E. Malone, Laura Muhlenbruck, Kelly Charlton and Harris Cooper. 2003. "Cues to Deception." *Psychological Bulletin* 129(1):74–118. S20

Ekman, Paul, Maureen O'Sullivan, Wallace V. Friesen and Klaus R. Scherer. 1991. "Invited Article: Face, Voice, and Body in Detecting Deceit." *Journal of Nonverbal Behavior* 15(2):125–135. S20

Elkins, Aaron, Stefanos Zafeiriou, Maja Pantic and Judee Burgoon. 2014. Unobtrusive Deception Detection. In *The Oxford Handbook of Affective Computing*, ed. Rafael A. Calvo, Sidney D'Mello, Johnathan Gratch and Arvid Kappas. New York, NY: Oxford University Press chapter 38, pp. 503–515. S20

Giannakopoulos, Theodoros and Aggelos Pikrakis. 2014. *Introduction to audio analysis: a MATLAB® approach*. Academic Press. S2, S27, S28

Goudbeek, Martijn and Klaus Scherer. 2010. "Beyond arousal: Valence and potency/control cues in the vocal expression of emotion." *The Journal of the Acoustical Society of America* 128(3):1322–1336. S2, S21, S22, S24

Grimmer, Justin and Brandon M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21(3):267–297. S29

Heilman, Kenneth M., Robert T. Watson and Edward Valenstein. 2011. Neglect and Related Disorders. In *Clinical Neuropsychology*, ed. Kenneth M. Heilman and Edward Valenstein. New York, NY: Oxford University Press pp. 296–348. S19

Hess, Wolfgang J. 2007. Pitch and Voicing Determination of Speech with an Extension Toward Music Signal. In *Springer Handbook of Speech Processing*, ed. Jacob Benesty, M. Mohan Sondhi and Yiteng Huang. 1st ed. Springer chapter 10, pp. 181–211. S5

Jones, Barbara E. 2003. "Arousal Systems." *Frontiers in Bioscience* 8:438–451. S19

Keele, Luke John. 2008. *Semiparametric Regression for the Social Sciences*. John Wiley & Sons. S42

Kim, Changick and Jenq-Neng Hwang. 2002. "Object-based video abstraction for video surveillance systems." *IEEE Transactions on Circuits and Systems for Video Technology* 12(12):1128–1138. S8

Klofstad, Casey A. 2016. "Candidate voice pitch influences election outcomes." *Political Psychology* 37(5):725–738. S20, S29

Koch, Michael and Ulrich Ebert. 1993. "Enhancement of the Acoustic Startle Response by Stimulation of an Excitatory Pathway from the Central Amygdala/Basal Nucleus of Meynert to the Pontine Reticular Formation." *Experimental Brain Research* 93(2):231–241. S19

Lang, Peter J., Mark K. Greenwald, Margaret M. Bradley and Alfons O. Hamm. 1993. "Looking at Pictures: Affective, Facial, Visceral, and Behavioral Reactions." *Psychophysiology* 30(3):261–273. S19

Laukka, Petri. 2004. Vocal expression of emotion: discrete-emotions and dimensional accounts PhD thesis Acta Universitatis Upsaliensis. S2, S23, S25

Maltzman, Forrest and Lee Sigelman. 1996. "The Politics of Talk: Unconstrained Floor Time in the U.S. House of Representatives." *The Journal of Politics* 58(3):819–830. S35

Mauss, Iris B. and Michael D. Robinson. 2009. "Measures of Emotion: A Review." *Cognition and Emotion* 23(2):209–237. S19

Morris, Jonathan S. 2001. "Reexamining the Politics of Talk: Partisan Rhetoric in the 104th House." *Legislative Studies Quarterly* 26(1):101–121. S32

Nadeau, David and Satoshi Sekine. 2007. "A survey of named entity recognition and classification." *Lingvisticae Investigationes* 30(1):3–26. S12

Nielsen, Finn Årup. 2011. "A new ANEW: Evaluation of a word list for sentiment analysis in microblogs." *arXiv preprint arXiv:1103.2903* . S15

Petrocik, John R. 1996. "Issue Ownership in Presidential Elections, with a 1980 Case Study." *American Journal of Political Science* 40(3):825–850. S42

Pisanski, Katarzyna, Judyta Nowak and Piotr Sorokowski. 2016. "Individual Differences in Cortisol Stress Response Predict Increases in Voice Pitch During Exam Stress." *Physiology & Behavior* 163:234–238. S19

Poole, Keith T. and Howard Rosenthal. 2001. "D-Nominate after 10 Years: A Comparative Update to Congress." *Legislative Studies Quarterly* pp. 5–29. S41

Posner, Jonathan, James A. Russell and Bradley S. Peterson. 2005. "The Circumplex Model of Affect: An Integrative Approach to Affective Neuroscience, Cognitive development, and Psychopathology." *Development and Psychopathology* 17:715–734. S19

Puts, David Andrew, Steven J.C. Gaulin and Katherine Verdolini. 2006. "Dominance and the Evolution of Sexual Dimorphism in Human Voice Pitch." *Evolution and Human Behavior* 27:283–296. S20

Roberts, Margaret E., Brandon M. Stewart and Dustin Tingley. 2014. "stm: R package for structural topic models." *R Package* 1:1–12. S41, S42

Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson and David G. Rand. 2014. "Structural Topic Models for Open-Ended Survey Responses." *American Journal of Political Science* 58(4):1064–1082. S41

Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley and Edoardo M. Airoldi. 2013. The Structural Topic Model and Applied Social Science. In *Advances in Neural Information Processing Systems Workshop on Topic Models: Computation, Application, and Evaluation*. S41

Roberts, Margaret E., Brandon M. Stewart and Edoardo M. Airoldi. 2016. "A Model of Text for Experimentation in the Social Sciences." *Journal of the American Statistical Association* pp. 1–49. S42

Rosen, Jeffery B., Janice M. Hitchcock, Catherine B. Sananes, Mindy J. D. Miserendino and Michael Davis. 1991. "A Direct Projection from the Central Nucleus of the Amygdala to the Acoustic Startle Pathway: Anterograde and Retrograde Tracing Studies." *Behavioral Neuroscience* 105(6):817–825. S19

Russell, James A. 1980. "A Circumplex Model of Affect." *Journal of Personality and Social Psychology* 39(6):1161. S19

Russell, James A. 2003. "Core Affect and the Psychological Construction of Emotion." *Psychological Review* 110:145–172. S19, S21, S27

Sheena, CV and NK Narayanan. 2015. "Key-frame extraction by analysis of histograms of video frames using statistical methods." *Procedia Computer Science* 70:36–40. S8

Zhuang, Yueting, Yong Rui, Thomas S Huang and Sharad Mehrotra. 1998. Adaptive key frame extraction using unsupervised clustering. In *Image Processing, 1998. ICIP 98. Proceedings. 1998 International Conference on*. Vol. 1 IEEE pp. 866–870. S8

Zuckerman, Miron and Robert E. Driver. 1985. Telling Lies: Verbal and Nonverbal Correlates of Deception. In *Multichannel Integrations of Nonverbal Behavior*, ed. Aaron W. Siegman and Stanley Feldstein. New York, NY: Lawrence Erlbaum chapter 3, pp. 129–147. S20