Supporting Information (to go online) for: Emotional Arousal Predicts Voting on the Supreme Court

Description of Supreme Court Audio Data

Although the U.S. National Archives maintain the official audio recordings, for this paper we accessed the audio files in .mp3 format via the Oyez Project at the Chicago-Kent College of Law, which is a public repository that aims "to be a complete and authoritative source for all audio recorded in the Court since the installation of a recording system in October 1955." To get additional information on the content of the files and to verify the identity of each speaker, we also obtained the transcript for each argument which can be found in the Oyez Project player.

As explained on Oyez[1], we do not have audio for every case. For example, some of the audio from the 1970s and 1990s resides in the National Archives in tape format. Even though Oyez is actively transferring this to a digital format, they still have not completed this process. There was also a problem with the Court's tape machine, especially during the early 1990s, which made the Justices talk slightly faster than normal. In order to synchronize these recordings with the transcripts, the Oyez team (especially Pat Ward) had to manually re-engineer the audio to make it "faithful to the speakers' voices."

Other problems arose in the 1980s. Unbeknownst to the Court, their tape manufacturer changed the formula they used to create the reel-to-reel tapes used for oral arguments. Instead of lasting 80 years, like typical reel-to-reel tapes, these new tapes only lasted a few years. Ultimately, they were plagued by what Oyez calls the "sticky shed syndrome" in which tape reels tended to stick together when stored. To solve this problem, the Oyez team had to bake the tapes in a slow oven for several hours, then mount the tapes and dub them to another real or digital storage device. Unfortunately, this was only partially successful and many of the recordings from the 1980s were lost.

These and other problems described in the "About the audio" portion of the Oyez website means the majority of our dataset is from 1998-2014. As these are essentially random errors,

[1]https://www.oyez.org/about

we use the full dataset in all of our models. For example, the Justices were unaware that the manufacturer had changed the composition of the reel-to-reel tapes. Similarly, we are also confident the Justices did not change their behavior based on a recording error that was uncovered decades later. Moreover, our results are robust to excluding years where we have a number of missing cases (i.e., 1980s and early 1990s). Indeed, our results remain the same regardless of whether our models are estimated using data from 1981-2014 or 1998-2014. Below we provide additional detail about the data collection process.[2]

Creating Statements from Utterances

Since the Oyez transcripts return a line for every utterance, we first collapsed all the utterances into "statements." For example, if Justice Roberts said, "I believe your argument is incorrect. Your client is guilty," Oyez splits this statement into two utterances. Specifically, they report one line for "I believe your argument is incorrect," and another line for "Your client is guilty." This is due to the nature of the "transcript," which is actually the scrolling text from the Oyez player. An example of this player can be found at this URL:

`https://web.archive.org/web/20150701205416/http://www.oyez.org/cases/2010-2019/2012/2012_12_399/argument`

Extracting the Audio

Once we obtained "statements," we extracted the audio. This was done using *ffmpeg*, a command line program for audio and video processing. Using the `R` statistical software language, we created a .csv file that can be fed directly into *ffmpeg* using a simple bash script. This script can be found here:

`author_website/code/split_transcript.sh`

Ultimately, this script will return one .wav file for each statement. These were saved in a separate folder called "utterances." This step is, by far, the most time consuming. Even using multiple processors on Amazon's Web Services (AWS), splitting the audio files took several weeks. Examples of the resulting output can be found here:

`author_website/audio/2012_11-626_argument_s53.885_s65.257.wav`

`author_website/audio/2012_11-626_argument_s71.494_s81.182.wav`

[2] Our data collection took place between June 1, 2015 and August 1, 2015. Sometime between September 6, 2015 and October 12, 2105 Oyez changed their website dramatically. The new version of the website contains more aggressive web scraping barriers. We provide links to a version of the website captured on July 1, 2015, even though we acknowledge this is not reflective of the current site. We chose this version of the website because it is reflective of the website we worked with in order to collect our data.

Extracting the Vocal Pitch

With these .wav files in hand, we used *Praat*[3] to extract vocal pitch. This commonly used software implements the algorithm outlined by Boersma (1993). Similar to other algorithms which focus on time-domain periodicity, *Praat* estimates the fundamental frequency by dividing the autocorrelation of a windowed signal by the autocorrelation of the window itself. One must assume the signal is stationary within each window, which is why the algorithm divides the audio file into small segments (around 60ms), then takes the average.[4] To ensure reliability, we excluded any questions and non-questions that were less than one second. Examples of the resulting output can be found here:

`author_website/pitch/2012_11-626_argument_s53_885_s65_257.txt`

`author_website/pitch/2012_11-626_argument_s71_494_s81_182.txt`

For those interested in extracting vocal pitch using *Praat* in batch, you can find the script we used here:

`author_website/code/extract_pitch.txt`

To use this script, one has to create two directories, "audio" and "pitch." Once these directories are created, place the `.wav` files of interest in the "audio" directory, then open *Praat* and click on "Open Praat Script."

Adding Vocal Pitch to the Transcripts

Using he `R` statistical software language, we added the pitch results to the transcripts, which were collapsed into statements. If we had vocal pitch for a given statement, it is recorded, otherwise there is an NA. Statements without vocal pitch were not included in any of our analyses. An example of the transcripts without vocal pitch can be found here:

`author_website/initial_results/transcript_example.csv`

An example of the final results (including vocal pitch) can be found here:

`author_website/initial_results/results_example.csv`

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

[4] Specifically, to use this software, one has to set five parameters: the pitch floor, pitch ceiling, window length, window shape, and voicing threshold. We set the pitch floor and ceiling to 50Hz and 300Hz, respectively. This resulted in a window length of 60ms. For both the window shape and voicing threshold we used the default settings

Adding Other Features from Oyez

Once the vocal pitch was added, we added a few features to the results in order to make later calculations possible. First, we added each speaker's "type." For each transcript, Oyez indicates whether the speaker is a "justice," "advocate," "unidentified," or "other." We used these classifications to determine whether a Justice or attorney ("advocate") was speaking. Second, we also added the description for each advocate. This helped us identify the petitioner and respondent. To understand this step, one should reference this URL:

```
https://web.archive.org/web/20150701205416/http://www.oyez.org/cases/2010-2019/
2012/2012_12_399
```

Notice under "Lisa S. Blatt" it says "for the petitioner." Similarly, under "Paul D. Clement" it says "for respondent Guardian ad Litem in support of the petitioner." Using the following `R` code, we added these statements and speaker types to the results:

```
author_website/code/get_advocates.R
```

Getting Each Justices' Baseline

Next, we had to find each Justices' overall mean and standard deviation. Women typically speak at a higher vocal pitch than men. Given that, we created baseline measures for each Justice using all of the audio from their questions and non-questions. These can be found in Table S1. Using these measures, we re-scaled vocal pitch to standard deviations above and below each Justices' average vocal pitch. For example, Justice Kagan's mean vocal pitch is 171.60Hz with a standard deviation of 28.05Hz. If she asked a question with a vocal pitch of 200Hz, our re-scaled measure would be $\frac{200-171.60}{28.05} = 1.01$, suggesting for that question her vocal pitch was a little over one standard deviation higher than her baseline.

For this step, we had to do some adjustments since Oyez uses multiple names for some Justices. For example, John Roberts is recorded as either "john_g_roberts" or "john_g_roberts_jr." Once these adjustments were made, we created baseline measures using every audio file available for each Justice. To replicate this step, download the following .zip file:

```
author_website/code/justice_baseline.zip
```

Inside the .zip file, you should find a sample dataset and code. For those who wish to fully replicate this step, the full version of the dataset is available upon request.

Getting the Vocal Pitch Results

With these variables added, we compiled the vocal pitch results necessary for our paper. In essence, we cycle through each line of a transcript and ask whether the statement is longer than a second. When a statement met this criteria, we record whether it contained a question mark, then we obtained the baseline measure for the Justice speaking. Using these baseline measures, we converted the given vocal pitch into standard deviations above or below the Justices' baseline. Once this is done, we look for an attorney in the lines following the question/statement. If one is found, we determine whether the attorney is a "respondent" or "petitioner," using the information outlined above. This process was repeated for questions and non-questions at both the case and Justice-level. To replicate this step, download the following .zip file:

```
author_website/code/pitch_results.zip
```

Inside the .zip file, you should find a sample dataset and code. For those who wish to fully replicate this step, the full version of the dataset is available upon request.

Finalizing the Results

With the audio data collected, we added the vocal pitch results to the Justice-level Supreme Court Database. After that, we added measures of Justice ideology. In the next section, we describe this data in more detail. The data needed to replicate Table 1 in the main text can be found here:

```
author_website/final_results/justice_results.csv
```

Please use the following code to estimate the models we used in the paper:

```
author_website/code/final_models.R
```

Description of Non-Audio Supreme Court Data

As described above, we linked the audio recordings from Oyez to data on Supreme Court rulings, specifically which party (the petitioner versus the respondent) won or lost each case and how each Justice eventually voted. For this, we relied on the existing Supreme Court Database (Spaeth et al. 2015), which provides case- and Justice-level data on the identities of the petitioner and respondent, what the case was about, and how the Justices voted (for or against the petitioner or respondent). We linked these to the oral arguments audio data by using the docket number of each case.

We have several analyses in which we control or analyze Justice ideology. Although there are a variety of ways to measure ideology on the Supreme Court (Schubert 1965, 1974; Rohde and Spaeth 1976; Segal and Cover 1989), we use scores developed by (Martin and Quinn 2002). The "Martin-Quinn" scores are estimated using dynamic item response theory, which allows Justice ideology to be estimated for every Justice serving from 1937 to 2014. Since their publication, these scores have been the primary way to measure ideology on the Supreme Court, despite a small number of scholars who question their utility in some instances (Farnsworth 2007). For a reply, please consult (Epstein et al. 2007). The Martin-Quinn scores are a continuous measure from conservative to liberal and have been substantially cross-validated with other measures.

In addition to ideology, we included the controls used by Black et al. (2011). The `Number More Questions Directed at Petitioner` is the number of questions asked of the petitioner minus the number of questions asked of the respondent. `Lower Court Decision Was Conservative` is a binary variable where a 1 indicates the decision of the court being reviewed was ideologically conservative and 0 if it was liberal. Since litigants may benefit from the federal government's support (Segal 1988), we included `Solicitor General as Amicus Supports Petitioner` and `Solicitor General as Amicus Supports Respondent` which returns 1 when the federal government supports the petitioner and respondent, respectively. Similarly, when litigants garner more support from interest groups, they are more likely to win (Collins 2004, 2008; Songer and Sheehan 1993). To control for this effect, we included `Number of Amicus Briefs Supporting Petitioner` and `Number of Amicus Briefs Supporting Respondent` which are the total number of amicus briefs filed on behalf of the petitioner and respondent, respectively. Finally, we used the Supreme Court database party codes to capture the petitioner's and respondent's level of resources. Similar to Black et al. (2011) and Collins (2004, 2008) we placed the petitioners and respondents into the following categories: (1) poor individuals, (2) minorities, (3) individuals, (4) unions or interest groups, (5) small businesses, (6) businesses, (7) corporations, (8) local governments, (9) state governments, and the U.S. government (10). In these categories, we assume poor individuals (1) have the least resources and the U.S. government (10) have the most resources. For more details, please see Black et al. (2011, 576).

"Positive" and "Negative" Words

We used three dictionaries to assess the degree to which Justices used "positive" and "negative" words during oral arguments. 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:

acceptable

benevolent

charming

devoted

elegant

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:

abusive

brutal

contempt

destructive

envious

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

beloved

compassionate

desirable

enjoyable

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

abrasive

brutish

condescending

derisive

exterminate

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:

admirable

beautiful

cheerful

delighted

enthusiastic

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

alarming

bitterness

cruelty

disgusting

enrage

Undoubtedly, our version of DAL is not ideal, but it is the only version we could obtain. Given that it is a modified version, we provide all the "positive" and "negative" words we used from DAL:

`author_website/dictionaries/dal.csv`

Unfortunately, we can not provide the LIWC categories. However, to give some comparison, we provide all the "positive" and "negative" words we used from the Harvard IV-4 dictionary:

`author_website/dictionaries/harvard.csv`

Descriptive Statistics

Below we report several descriptive statistics. We begin with each Justices' baseline measures. These are reported in Table S1. Of the female Justices, Sandra Day O'Connor had the highest average vocal pitch (196.28Hz) and had the highest standard deviation (31.87Hz). For the male Justices, Thurgood Marshall had the highest average vocal pitch (147.84Hz), whereas Harry Blackmun had the highest standard deviation (46.96Hz).

On average, each question and non-question was 15 and 21 seconds long, respectively. William Burger asked the longest question which was 20 minutes long. The shortest question, asked by Antonin Scalia, was less than a second. For non-questions, Antonin Scalia gave both the longest and shortest. The former was close to four minutes long. The latter was less than a second. For reasons explained below, these incredibly short questions and

Table S1: Average vocal pitch and standard deviation for U.S. Supreme Court Justices, 1982–2014.

Justice	Pitch Mean	Pitch SD	Questions	Non-Questions	Total
Sandra Day O'Connor	196.28	31.87	3,505	3,115	6,620
Elena Kagan	171.60	28.05	1,616	1,363	2,979
Sonia Sotomayor	166.14	25.07	3,345	3,165	6,510
Ruth Bader Ginsburg	163.82	22.54	7,284	8,509	15,793
Thurgood Marshall	147.84	33.75	292	173	465
Antonin Scalia	136.57	31.97	12,382	15,523	27,905
David Souter	135.62	29.43	5,404	5,816	11,220
John Roberts	133.07	28.24	5,261	9,308	14,569
William Brennan	127.48	28.63	5	9	14
Stephen Breyer	126.95	30.62	9,094	10,643	19,737
Harry Blackmun	122.24	46.96	63	31	94
Anthony Kennedy	121.52	22.31	6,208	7,191	13,399
William Rehnquist	121.07	25.25	3,098	5,659	8,757
Warren Burger	119.65	33.49	539	1,279	1,818
Lewis Powell	117.01	36.97	78	42	120
Byron White	116.58	33.42	193	240	433
John Paul Stevens	116.21	24.74	5,505	5,727	11,232
Samuel Alito	112.66	24.67	3,366	1,239	4,605
Clarence Thomas	101.81	24.30	31	34	65
Average	134.43	29.59	3,540	4,161	7,702

Note: Measurements of vocal pitch are in Hertz (Hz). To calculate each we used all of the audio from Justices' questions and non-questions. The number of questions and non-questions are reported in the corresponding columns. In the last column, we report the total number of utterances. Averages of each column are reported at the bottom of the table. The table is sorted by the mean vocal pitch.

non-questions were excluded from the analysis, although they do not substantively affect the results.

Ultimately, we found William Brennan spoke the least, followed by Clarence Thomas. In total, Brennan spoke for a little under 3 minutes across 14 questions/non-questions. Clarence Thomas spoke for a little over 11 minutes across 65 question/non-questions. On average, Justices spoke a little over 25 hours across 7,702 question/non-question, suggesting Brennan and Thomas barely spoke.

Antonin Scalia was the most loquacious Justice, asking 12,382 questions and making 15,523 statements. In total, Scalia spoke for 79.82 hours, which is over three times as much as the average Justice. His next closest competitor, Stephen Breyer, spoke close to 7 hours longer than Scalia, but did so using 3,288 fewer questions and 4,880 fewer non-questions.

The differences found between Scalia and Brennan should be taken into consideration when evaluating our results. The same can be said for Breyer and Thomas.

Table S2: Average pitch difference and standard deviation for U.S. Supreme Court Justices, 1982–2014.

Justice	Pitch Difference Mean	Pitch Difference SD	Cases
John Roberts	0.05	0.59	180
Anthony Kennedy	0.02	0.49	571
Lewis Powell	0.01	0.18	2
Ruth Bader Ginsburg	−0.01	0.66	746
Stephen Breyer	−0.02	0.54	639
Thurgood Marshall	−0.02	0.58	12
John Paul Stevens	−0.03	0.42	428
David Souter	−0.03	0.50	421
Sonia Sotomayor	−0.04	0.53	225
Antonin Scalia	−0.09	0.57	774
Samuel Alito	−0.09	0.57	146
Elena Kagan	−0.10	0.53	149
Sandra Day O'Connor	−0.10	0.57	313
William Rehnquist	−0.14	0.65	274
Byron White	−0.14	1.34	9
Clarence Thomas	−0.47	–	1
Harry Blackmun	−0.68	1.22	4
Warren Burger	−0.80	1.75	44
Average	−0.15	0.69	246.90

Note: For each Justice, we converted vocal pitch to standard deviations above and below his or her average vocal pitch. The average vocal pitch (standardized) in questions directed towards the petitioner ("Petitioner Pitch") minus the average vocal pitch (standardized) in questions directed towards the respondent ("Respondent Pitch") is captured in "Pitch Difference" (Petitioner Pitch - Respondent Pitch). In the first and second columns, we report the mean and standard deviation of this measure for each Justice. We report the total number of cases in the last column. Averages of each column are reported at the bottom of the table. The table is sorted by the mean "Pitch Difference."

Table S2 reports the mean and standard deviation of "Pitch Difference" for each Justice. In this table, positive values imply the Justice generally spoke at a higher vocal pitch towards the petitioner. Unsurprisingly, all but three Justices tended to speak at a higher vocal pitch when addressing the respondent. As indicated in the main text, petitioners win the majority of cases, so one would expect Justices to speak at a higher vocal pitch towards the respondents. It is also worth noting most of the standard deviations are all less than one, suggesting we are capturing very small changes in vocal pitch.

We also checked to see if outliers were unduly affecting our results. We estimated three different versions of the models outlined in Table 1 to assuage these concerns. First, we used Cook's distances to identify potential outliers using a $\frac{4}{n}$ cut off. When these observations (66) were identified, we added a dummy variable indicating whether the case was an "outlier" and re-estimated Models (1), (2), (3), and (4). In each of these models "Pitch Difference" was still statistically significant at the 0.0001-level. Second, we re-estimated Models (1), (2), (3), and (4) using robust logistic regression. Even though the `R` statistical software language has two packages that estimate robust logistic regressions, we used the `robust` package. Again, in each of these models "Pitch Difference" was statistically significant at the 0.0001-level. Finally, we re-estimated Models (1), (2), (3), and (4) using standard errors from 10,000 bootstrapped samples. The 95 percent confidence intervals for "Pitch Difference" in each of these models were [-0.15,-0.36], [-0.15,-0.31], [-0.15,-0.31], and [-0.15,-0.31], respectively. None of these confidence intervals overlap zero, suggesting we still find a statistically significant relationship even when the standard errors are bootstrapped.

Cross-Validation

Creating cross-validated predictions from a multilevel model is not entirely straightforward. If one were to simply divide the data into k equally sized folds, then there is a chance that some Justices would not appear in every fold. One way to resolve this issue is to randomly leave one Justice out, then use the other Justices to predict the left out Justice's votes. Although this approach makes intuitive sense when there are a large number of higher level groups (i.e., states or countries), when one only has a handful of Justices this would not only exclude large portions of data, but it would also create folds of varying size. In k-fold cross-validation we are interested in obtaining the prediction rate out of sample, both of the aforementioned problems would impede our ability to achieve this end. Moreover, such an approach implies one Justice's vocal inflections patterns can be used to predict another's which defeats the whole purpose of using a multilevel model.

With this in mind, we created k equally sized folds for each Justice. Beginning with Antonin Scalia, we restrict the data to only rows in which the selected Justice was involved. Once subsetted, we then randomly divide those rows into k equally sized folds (with replacement). This process was repeated for each Justice until we had a dataset in which each fold had approximately the same proportion dedicated to each Justice. For example, Justice Scalia accounted for around 18 percent of the questions asked. This means he would represent approximately 18 percent of each fold. Unfortunately, there is no definitive way to create a cross-validation sample from a multilevel dataset, but we think our approach is reasonable. We encourage those interested to view our replication code:

```
author_website/code/cross_validation.R
```

Individual Justice Analysis

Since some Justices are more active on the bench than others, we also estimated another multilevel logistic regression in which we randomly varied “Pitch Difference" by Justice. Substantively, this model tests whether the predictive power of vocal pitch is restricted to only a handful of Justices. In Table S3, the randomly varying intercept was said to be “significant" ($p < .05$) if a simple ANOVA suggested it explained more of the variance as compared to a simple logistic regression without a randomly varying intercept. Similarly, the randomly varying slope was said to be “significant" ($p < .05$) if it was found to explain more of the variance than a multilevel logistic regression with a randomly varying intercept. Regardless of the model specification, randomly varying “Pitch Difference" did not significantly improve the performance of the model, suggesting the predictive power of vocal pitch does not vary considerably from one Justice to the next.

Figure S1 shows bootstrapped 95 percent confidence intervals for the coefficient estimated for each Justice in Model 1. If any of the confidence intervals overlap zero, then it would suggest vocal pitch does not help explain the corresponding Justice’s votes. With the exceptions of Justices Sotomayor and Blackmun, all the random coefficients are negative, suggesting that for 16 of 18 Justices an increase in vocal pitch is predictive of a vote against the lawyer the Justice is addressing. Moreover, the confidence intervals for 12 of 16 Justices do not overlap zero, suggesting the predictive power of vocal pitch is statistically significant for the vast majority of Justices. Although not shown here, the random coefficients for Justice Roberts, Rehnquist, Burger, and Marshall are statistically significant at the .10-level, suggesting for most Justices vocal pitch carries some statistical weight.

Figure S2 shows the predicted probabilities from the six Justices who appear most in our data. These are the Scalia (27,905 utterances), Stephen Breyer (19,737), Ruth Bader Ginsburg (15,793), frequent “swing vote” Anthony Kennedy (13,399), John Paul Stevens (12,232), and David Souter (11,220). These probabilities are derived using the coefficients from Model 1 in able S3. For all Justices, there is a negative relationship: when the vocal pitch in questions directed to the petitioner is greater (suggesting greater arousal), the petitioner is more likely to lose that Justice’s vote. To put this into perspective, Justice Scalia spoke with an average vocal pitch of 136.57Hz (in layman’s terms roughly equivalent to slight skepticism),[5] with a standard deviation of 31.97Hz. If he had spoken at 168.54Hz (stronger skepticism)[6] towards the petitioner (one standard deviation above his baseline) and 104.60Hz (neutral tone)[7] towards the respondent (one standard deviation below his baseline), then the predicted probability of the petitioner winning his vote would be 34.02 percent, or 18.60 percentage points lower than would be expected if his vocal pitch was the same towards the petitioner and respondent (52.50 percent). The magnitude of vocal pitch’s predictive power is similar for Justice Anthony Kennedy, a frequent swing vote in controversial 5–

[5] See an example at `www.project-website.com/audio/scalia_average_pitch.wav`.

[6] For example, `www.project-website.com/audio/scalia_high_pitch.wav`.

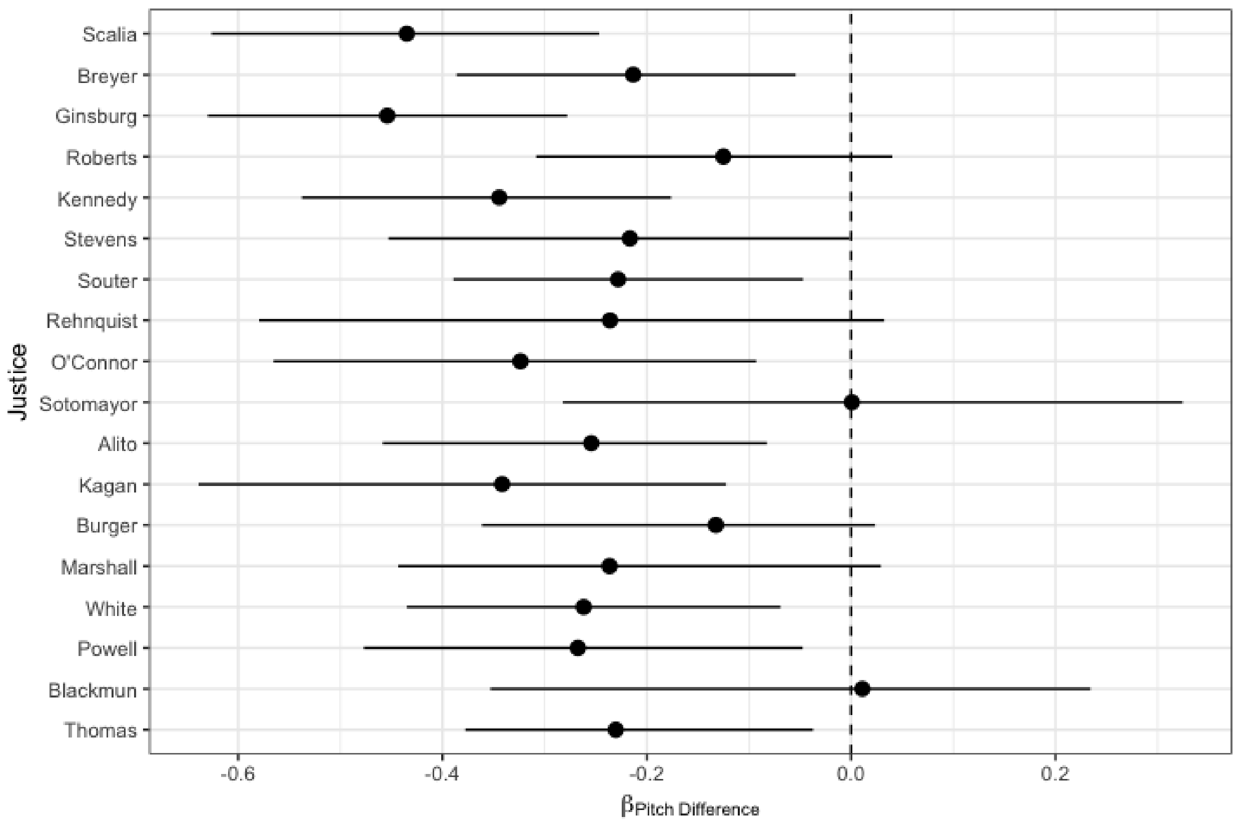
[7] For example, `www.project-website.com/audio/scalia_low_pitch.wav`.

Table S3: Does Vocal Pitch Predict Votes in Favor of the Petitioner? (Random Slope)

	No Controls	DAL	Harvard IV	LIWC
	(1)	(2)	(3)	(4)
Fixed Effects				
Constant	0.185***	−0.025	−0.027	−0.026
	(0.056)	(0.161)	(0.161)	(0.161)
Pitch Difference	−0.248***	−0.206***	−0.206***	−0.206***
	(0.050)	(0.049)	(0.049)	(0.049)
Percent More Unpleasant Words Directed at Petitioner		−2.011	0.080	−2.179*
		(1.466)	(0.848)	(1.314)
Percent More Pleasant Words Directed at Petitioner		−1.631	0.285	−1.666
		(1.083)	(0.685)	(1.047)
Number More Questions Directed at Petitioner		−0.056***	−0.056***	−0.056***
		(0.008)	(0.008)	(0.008)
Political Ideology$_{t-1}$		0.158***	0.159***	0.158***
		(0.033)	(0.033)	(0.033)
Lower Court Decision Was Conservative		0.013	0.014	0.014
		(0.073)	(0.073)	(0.073)
Political Ideology$_{t-1}$ × Lower Court Decision Was Conservative		−0.263***	−0.263***	−0.263***
		(0.034)	(0.034)	(0.034)
Solicitor General as Amicus Supporting Petitioner		0.543***	0.546***	0.543***
		(0.079)	(0.079)	(0.079)
Solicitor General as Amicus Supporting Respondent		−0.671***	−0.665***	−0.671***
		(0.104)	(0.104)	(0.104)
Number of Amicus Briefs Supporting Petitioner		0.039***	0.039***	0.039***
		(0.008)	(0.008)	(0.008)
Number of Amicus Briefs Supporting Respondent		−0.058***	−0.059***	−0.058***
		(0.007)	(0.007)	(0.007)
Petitioner's Level of Resources		0.045***	0.046***	0.045***
		(0.014)	(0.014)	(0.014)
Respondent's Level of Resources		−0.003	−0.003	−0.003
		(0.014)	(0.014)	(0.014)
Random Effects				
Intercept	0.03***	0.02**	0.02**	0.02**
	(0.16)	(0.13)	(0.12)	(0.12)
Pitch Difference	0.02	0.01	0.01	0.01
	(0.12)	(0.11)	(0.11)	(0.11)
N_1	5,209	4,977	4,977	4,977
N_2	18	18	18	18
$\log L$	−3,549.531	−3,200.175	−3,202.095	−3,199.672
AIC	7,109.061	6,434.350	6,438.189	6,433.344
Percent Correctly Predicted	55.72	62.78	62.63	62.84

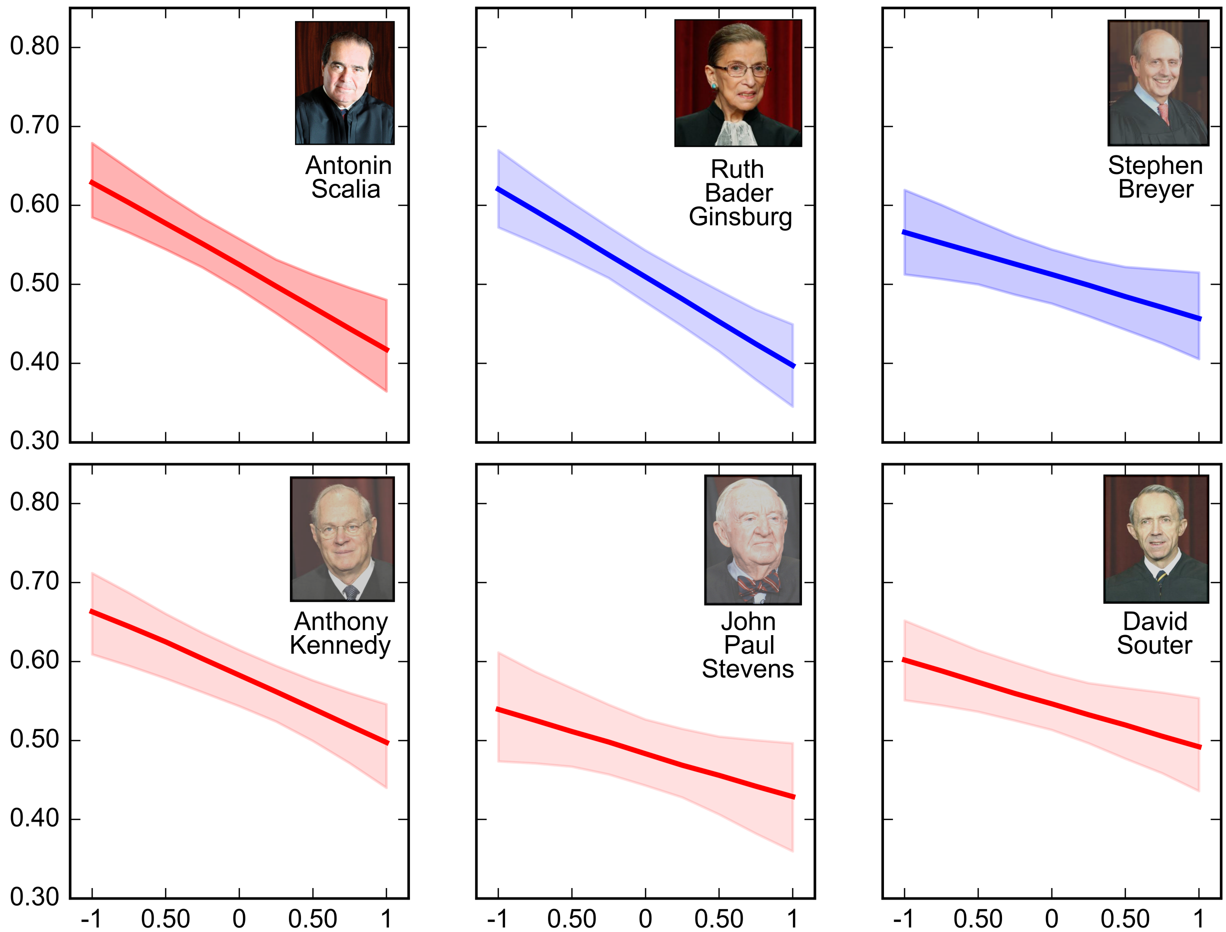
Note: Each model is a multilevel logistic regression with a random intercept for each Justice. Outcome is whether the Justice voted in favor of petitioner. Unit of analysis is each Justice's vote. Models include statements with question marks. The average vocal pitch in questions directed towards the petitioner ("Petitioner Pitch") minus the average vocal pitch in questions directed towards the respondent ("Respondent Pitch") is captured in "Pitch Difference" (Petitioner Pitch - Respondent Pitch). Model 2 uses the Dictionary of Affect in Language (DAL). Model 3 uses the Harvard-IV dictionary. Model 4 uses the Linguistic Inquiry and Word Count (LIWC) dictionary. The rest of the controls are the same as Black et al. (2011). 10-fold cross-validation is used to estimate the percent of votes correctly predicted. Please refer to the Supplemental Information for more details about each dictionary, the controls, and our cross-validation approach. Levels of significance are reported as follows: *p < .1; **p < .05; ***p < .01.

Figure S1: Random Slopes for Each Justice Using Model 1 from Table S3



Note: This plot shows bootstrapped 95 percent confidence intervals using the coefficients from the first model in Table S3. The bootstrapped samples were created by first randomly sampling (with replacement) from the Justices, then randomly sampling (with replacement) from within each Justice's subset of the data. The dots represent the median of the bootstrapped sample. The lines represent the 95th percentile. If the line overlaps zero (see dashed line), then the corresponding random coefficient is not statistically significant.

Figure S2: Does Vocal Pitch Predict Justices' Votes? (Random Slope)



Note: Plotting predicted probabilities for the six most frequently speaking Justices. Outcome is whether the Justice voted in favor of petitioner. Unit of analysis is each Justice's vote. Both fixed and random coefficients are from Model 1 in Table S3. The only independent variable is "Pitch Difference" (the average vocal pitch in questions directed towards the respondent from the average vocal pitch in questions directed towards the petitioner). The x-axis is "Pitch Difference" from -1 to 1. The y-axis is the probability the Justice voted for the petitioner. 95 percent confidence intervals are included. Darker ribbons imply the Justice appeared more frequently in the data. Darker pictures imply the Justice spoke at a higher vocal pitch. Ruth Bader Ginsburg spoke at an average vocal pitch of 163.82Hz. John Paul Stevens spoke at an average vocal pitch of 116.21Hz. The plots are organized from the Justice that appeared in our data the most (Scalia) to the least (Souter). Red indicates the Justice was appointed by a Republican president. Blue indicates the Justice was appointed by a Democratic president.

4 cases. When Justice Kennedy speaks one standard deviation (22.31Hz) higher than his baseline (121.52Hz)[8] towards the petitioner (143.83Hz)[9] and speaks to the respondent one standard deviation below his baseline towards the respondent (99.21Hz),[10] the probability Justice Kennedy votes for the petitioner is 42.77 percent. When Justice Kennedy speaks at the same vocal pitch to both the petitioner and respondent, the probability he votes for petitioner is 15.33 percentage points higher (or 57.90 percent). Regardless of the Justice, vocal pitch is highly predictive of their eventual vote, suggesting even when Justices display subtle emotion, such expressions are highly informative of their underlying predispositions.

Ninth Circuit Analysis

We also recently applied our model to the 9th Circuit Court of Appeals. When the Trump Administration issued an executive order that banned entry from seven majority-Muslim countries, several legal challenges were immediately filed. Within two weeks, these challenges reached the 9th Circuit, who heard oral arguments on February 7, 2017. More than 130,000 people listened to the arguments, and hundreds of experts weighed in on how the judges would vote. We saw this as a good opportunity to see whether vocal pitch could be used to predict cases other than those found on the Supreme Court.

The oral arguments began at 7:09PM EST. Within the hour we had downloaded the one hour and eight minutes of audio from the 9th Circuit's website. Using Audacity,[11] we then split the audio into individual speech acts, focusing exclusively on Judges Canby, Clifton, and Friedland. This process took approximately three and half hours. With the segmented audio in hand, we then extracted the vocal pitch using Praat.[12] At 11:59PM EST, we posted the results shown in Figure S3. Based on the results found in the present paper, we were confident the state of Washington was going to win its case.

First, the vocal pitch was significantly higher ($t = 2.83$, df = 75, $p < 0.001$) in questions directed towards the DOJ attorney (203.26Hz) as compared to questions directed towards Washington's Attorney General (180.53Hz). Part of this could be due to Judge Friedland. As mentioned in the main text, women tend to speak at a higher vocal pitch than men, meaning Judge Friedland could naturally raise the vocal pitch of questions directed towards the DOJ attorney by simply participating more. In the present paper, we deal with this issue by first standardizing vocal pitch to standard deviations above or below a speaker's mean vocal pitch. Unfortunately, we do not have years worth of audio data to create baselines for each judge. With that said, when the same calculation is done using only the male judges we found the same result – vocal pitch was higher in questions directed towards the DOJ attorney (168.99Hz) as compared to questions directed towards Washington's Attorney Gen-

[8] See an example at `www.project-website.com/audio/kennedy_average_pitch.wav`.

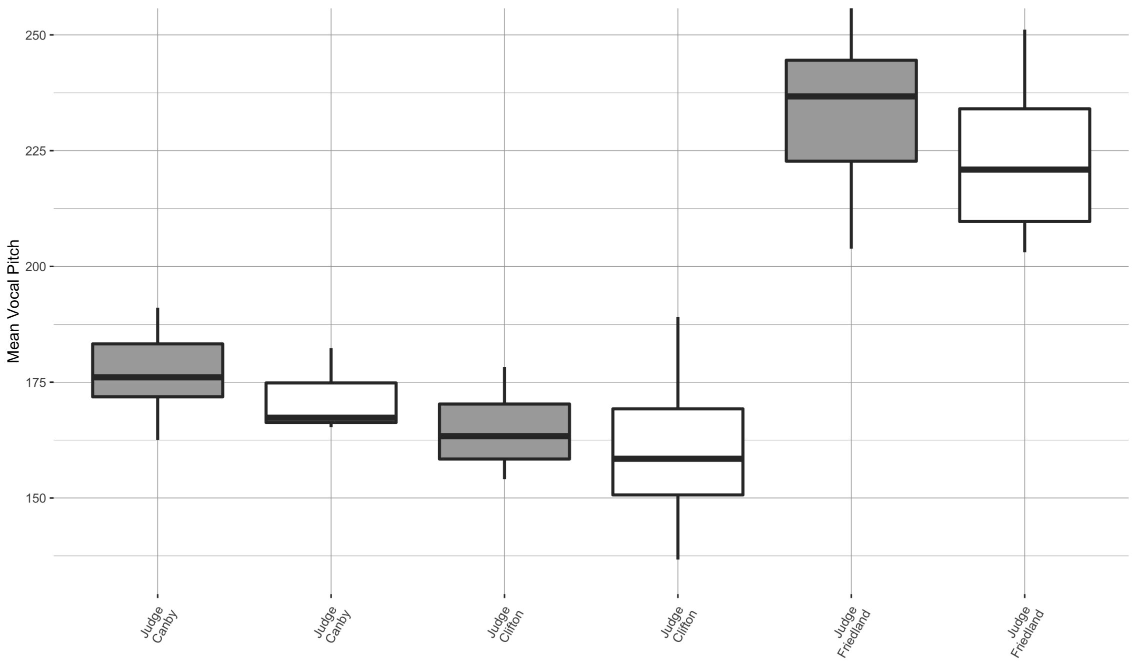
[9] For example, `www.project-website.com/audio/kennedy_high_pitch.wav`.

[10] For example, `www.project-website.com/audio/kennedy_low_pitch.wav`.

[11] `http://www.audacityteam.org`

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

Figure S3: Average Vocal Pitch from Travel Ban Oral Arguments on February 7, 2017



Note: In this plot gray boxes capture the vocal pitch of questions directed towards the attorney representing the Department of Justice, whereas white boxes capture the vocal pitch of questions directed towards Washington's Attorney General. Moving from left to right, you will find box plots for Judge Canby, Clifton, and Friedland.

eral (161.78Hz). Even though this was only statistically significant at the .10-level ($t = 1.70$, df = 43, p < 0.10), we think this result suggests all judges tended to have an unfavorable disposition towards the DOJ attorney.

Second, we used the models estimated throughout the present paper to try to convert our general impression into a specific estimate. Again, this is difficult because we have on average a little over six minutes of audio from each 9th Circuit judge, whereas we have a little over 25 hours from each Justice. To put this into perspective, we have more audio from Clarence Thomas (11 minutes), than any of the 9th Circuit judges. Despite the small amount of audio, we were able to obtain reasonable justice and case-level predictions. To do so, we did the following:

1. We standardized vocal pitch to standard deviations above or below each judge's baseline. Unfortunately, these baselines were derived from a single case, meaning they are not as precise as the baselines estimated for the Supreme Court.

2. In general, "Pitch Difference" is the average vocal pitch (standardized) directed towards the petitioner minus the average vocal pitch (standardized) directed towards the respondent. In this case, the plaintiff was the state of Washington and the defendant was the United States.[13]

3. With this in mind, "Pitch Difference" is simply the average vocal pitch (standardized) in questions directed towards the Washington Attorney General minus the average vocal pitch (standardized) in questions directed towards the DOJ attorney.

4. For Judges Canby, Clifton, and Friedland this resulted in a "Pitch Difference" of -0.55, -0.32, and -0.29, suggesting each judge spoke at slightly lower (less than 1 standard deviation) vocal pitch towards the Washington Attorney General.

Using the coefficients from Model, Table 1, we estimated there was a 58.04, 56.53, and 56.37 percent chance Judges Canby, Clifton, and Friedland would vote *for* the Washington Attorney General. This suggested there was a strong probability of a 3-0 decision ruling against President Trump's travel ban. Our prediction was even stronger at the case-level. Here, the overall "Pitch Difference" was -0.30, suggesting on average the vocal pitch of the judges was about a third of a standard deviation lower towards the Washington Attorney General as compared to the DOJ attorney. When we used this to predict the case outcome using the coefficients from our case-level models outlined below, we estimated there was a 67.85 percent chance the Washington Attorney General would win his case.

In a unanimous decision on February 10, the 9th Circuit Court of Appeals ruled for the state of Washington, upholding the stay on President Trump's travel ban. Using just the vocal pitch, we predicted this decision three days in advance. We not only correctly estimated the direction, but we also estimated each judge's vote with some degree of certainty. These

[13]http://cdn.ca9.uscourts.gov/datastore/opinions/2017/02/09/17-35105.pdf

results do not imply we should only consider vocal pitch when making such predictions, but they do suggest vocal pitch has important statistical and substantive weight. In the next section, we provide further evidence of the importance of vocal pitch by comparing vocal pitch to the {Marshall}+ algorithm.

Using the {Marshall}+ Algorithm

We chose the {Marshall}+ algorithm as a baseline because they make their results readily available at this URL:

`https://github.com/mjbommar/scotus-predict`

For us, this was extraordinarily useful because it allowed us to get some sense of how well vocal pitch predicted Supreme Court outcomes. In the associated paper (Katz, Bommarito and Blackman 2014), the authors find their algorithm predicts 69.7 percent of cases and 70.9 percent of votes. We do not find the same is true for the cases and votes we considered (see Tables S4 and S5). Specifically, we took the cases and votes where we had vocal pitch measurements, then found those cases and votes in the {Marshall}+ results file. Once we did that, we counted the number of "correct" predictions as indicated by the authors. These are the results presented in our paper.

With that said, there are a couple of things to note. First, {Marshall}+ does not predict petitioner votes, rather they predict whether the Supreme Court affirmed/reversed the lower court decision. This is not the same thing as what we are predicting in our model. We have no sense of whether this would help or hinder the substantive interpretation of our results. From our experience with this study, we would suspect it is easier to predict the Court's opinion towards the lower court, but this is purely speculative.

Second, to make our predictions comparable to both Katz, Bommarito and Blackman (2014) and Martin et al. (2004) we use the present term as the testing set and use previous terms within the same natural court as the training set. In terms where the natural court changed (2005, 2009, and 2010) this became impossible since there were no previous years to use for training. In these years we used the same term as both the training and testing dataset. If we exclude these terms from Table S4 we correctly predict 65.88 percent of cases, which is 1.63 percentage points better than the {Marshall}+ algorithm (64.25 percent). If we exclude these terms from Table S5 we correctly predict 58.09 percent of votes whereas the {Marshall}+ algorithm correctly predicts 68.38 percent. In all four instances, the models improve significantly over chance.

Third, the Katz, Bommarito and Blackman (2014) end their analysis in 2012, so we will exclude 2013 and 2014 from our comparisons. Similarly, Oyez does not provide a lot of audio prior to 1998, so we will begin our analysis in 1998. If we use all the available years in both datasets (1981-2012) we correctly predict 65.91 percent of cases which is 1.15 percent

Table S4: Predicting Supreme Court Outcomes Using Vocal Pitch

(a) Case Outcomes

Term	Correctly predicted using Pitch Difference	χ^2
1998 (n = 58)	74.14	12.57 (p < 0.000)
1999 (n = 63)	60.32	2.29 (p < 0.131)
2000 (n = 65)	70.77	10.40 (p < 0.001)
2001 (n = 29)	65.52	2.21 (p < 0.137)
2002 (n = 69)	62.32	3.71 (p < 0.054)
2003 (n = 55)	74.55	12.29 (p < 0.000)
2004 (n = 69)	68.12	8.35 (p < 0.004)
2005 (n = 68)	70.59	10.72 (p < 0.001)
2006 (n = 54)	64.81	4.17 (p < 0.041)
2007 (n = 57)	63.16	3.44 (p < 0.064)
2008 (n = 57)	75.44	13.75 (p < 0.000)
2009 (n = 39)	76.92	10.26 (p < 0.001)
2010 (n = 56)	62.50	3.02 (p < 0.082)
2011 (n = 46)	50.00	0.00 (p < 1.000)
2012 (n = 55)	58.18	1.16 (p < 0.281)
Total (n = 840)	66.55	91.34 (p < 0.000)

(b) Justice Votes

Term	Correctly predicted using Pitch Difference	χ^2
1998 (n = 324)	53.40	1.36 (p < 0.243)
1999 (n = 336)	52.98	1.07 (p < 0.300)
2000 (n = 355)	54.65	2.88 (p < 0.089)
2001 (n = 124)	62.99	8.06 (p < 0.005)
2002 (n = 376)	56.65	6.39 (p < 0.012)
2003 (n = 309)	60.90	14.39 (p < 0.000)
2004 (n = 365)	59.73	13.42 (p < 0.000)
2005 (n = 353)	61.89	19.27 (p < 0.000)
2006 (n = 300)	57.76	6.98 (p < 0.008)
2007 (n = 305)	54.75	2.57 (p < 0.109)
2008 (n = 293)	66.89	32.78 (p < 0.000)
2009 (n = 206)	58.25	5.29 (p < 0.021)
2010 (n = 298)	53.69	1.48 (p < 0.224)
2011 (n = 267)	59.18	8.63 (p < 0.003)
2012 (n = 310)	51.94	0.39 (p < 0.532)
Total (n = 4521)	57.42	99.48 (p < 0.000)

Note: Percentage of correctly predicted Supreme Court outcomes. In the panel (a), the outcome is whether the petitioner won the case. In the panel (b), the outcome is whether the Justice voted in favor of petitioner. The second column reports the percentage of outcomes correctly predicted using only "Pitch Difference." For each, the testing dataset is the term listed in the table. The training dataset is all preceding terms within the same natural court. The χ^2 statistic tests the proportion of correctly predicted outcomes against the chance level of 50%.

Table S5: Comparing Vocal Pitch to the {Marshall}+ Algorithm

(a) Case Outcomes

Term	Correctly predicted using Pitch Difference	Correctly predicted using {Marshall}+	χ^2
1998 (n = 58)	74.14	67.24	0.37 (p < 0.541)
1999 (n = 63)	60.32	65.08	0.14 (p < 0.713)
2000 (n = 65)	70.77	63.08	0.56 (p < 0.456)
2001 (n = 29)	65.52	62.07	0.00 (p < 1.000)
2002 (n = 69)	62.32	65.22	0.03 (p < 0.859)
2003 (n = 55)	74.55	63.64	1.06 (p < 0.302)
2004 (n = 69)	68.12	68.12	0.00 (p < 1.000)
2005 (n = 68)	70.59	66.18	0.14 (p < 0.712)
2006 (n = 54)	64.81	68.52	0.04 (p < 0.838)
2007 (n = 57)	63.16	57.89	0.15 (p < 0.702)
2008 (n = 57)	75.44	82.46	0.48 (p < 0.491)
2009 (n = 39)	76.92	66.67	0.57 (p < 0.450)
2010 (n = 56)	62.50	67.86	0.16 (p < 0.592)
2011 (n = 46)	50.00	52.17	0.00 (p < 1.000)
2012 (n = 55)	58.18	50.91	0.33 (p < 0.566)
Total (n = 840)	66.55	64.76	0.52 (p < 0.472)

(b) Justice Votes

Term	Correctly predicted using Pitch Difference	Correctly predicted using {Marshall}+	χ^2
1998 (n = 324)	53.40	67.90	13.68 (p < 0.000)
1999 (n = 336)	52.98	70.54	21.20 (p < 0.000)
2000 (n = 355)	54.65	67.04	10.93 (p < 0.001)
2001 (n = 124)	62.99	70.16	1.14 (p < 0.285)
2002 (n = 376)	56.65	66.84	7.86 (p < 0.005)
2003 (n = 309)	60.90	69.58	4.78 (p < 0.029)
2004 (n = 365)	59.73	68.22	5.35 (p < 0.021)
2005 (n = 353)	61.89	66.95	1.75 (p < 0.186)
2006 (n = 300)	57.76	72.00	12.80 (p < 0.000)
2007 (n = 305)	54.75	69.51	13.49 (p < 0.000)
2008 (n = 293)	66.89	71.67	1.36 (p < 0.244)
2009 (n = 206)	58.25	62.14	0.50 (p < 0.481)
2010 (n = 298)	53.69	63.42	5.42 (p < 0.020)
2011 (n = 267)	59.18	65.92	2.31 (p < 0.129)
2012 (n = 310)	51.94	64.52	9.58 (p < 0.002)
Total (n = 4521)	57.42	67.79	103.33 (p < 0.000)

Note: Percentage of correctly predicted Supreme Court outcomes. In the panel (a), the outcome is whether the petitioner won the case. In the panel (b), the outcome is whether the Justice voted in favor of petitioner. The second column reports the percentage of outcomes correctly predicted using only "Pitch Difference." The third column reports the percentage of outcomes correctly predicted using the {Marshall}+ algorithm. For each, the testing dataset is the term listed in the table. The training dataset is all preceding terms within the same natural court. The χ^2 statistic tests the proportion of correctly predicted using "Pitch Difference" versus the {Marshall}+ algorithm.

higher than the {Marshall}+ algorithm (64.76 percent). If we use all the available years we correctly predict 57.68 percent of votes whereas the {Marshall}+ algorithm correctly predict 67.79 percent. In all four instances, the models improve significantly over chance.

Table S6: Controlling for the 30 Best {Marshall}+ Predictors

Model 1	
(Intercept)	13.42
	(9.67)
Pitch Difference	-0.32***
	(0.05)
Case Information	
Law Type	-0.02
	(0.02)
Lower Court Disposition	0.02
	(0.02)
Issue	0.00
	(0.00)
Issue Area	-0.74
	(2.86)
Month Argument	-0.00
	(0.01)
Month Decision	-0.04**
	(0.02)
Petitioner	-0.00
	(0.00)
Petitioner Binned	-0.03*
	(0.01)
Respondent	-0.00
	(0.00)
Respondent Binned	-0.02
	(0.01)
Cert Reason	0.02*
	(0.01)
Overall Historic Supreme Court Trends	
Mean Court Direction Issue	0.79*
	(0.33)
Mean Court Direction 10	-0.44
	(0.30)

Continued on next page

Table S6 – *Continued from previous page*

Lower Court Trends	
Mean Lower Court Direction Issue	0.38
	(0.26)
Current Supreme Court Trends	
Mean Current Court Direction Issue	-0.57*
	(0.29)
Std. Dev. Current Court Direction Circuit Origin	-2.04†
	(1.11)
Std. Dev. Current Court Direction Issue	0.15
	(0.42)
Individual Supreme Court Justice Trends	
Mean Justice Direction	-8.61
	(6.89)
Mean Justice Direction 10	0.06
	(0.31)
Mean Justice Direction Z-Score	33.57
	(27.98)
Difference in Trends	
Difference Court Lower Ct. Direction	-0.35
	(0.23)
Abs. Difference Court Lower Ct. Direction	-3.43†
	(1.98)
Z-Score Abs. Difference Court Lower Ct. Direction	-0.33*
	(0.15)
Difference Justice Court Direction	-0.12
	(5.68)
Difference Justice Court Direction Issue	-0.27
	(0.19)
Z-Score Justice Court Direction Difference	-0.08
	(0.56)
Agreement of Justice with Majority 10	1.17***
	(0.28)
N	3223
AIC	4231.16
BIC	4936.22
$\log L$	-1999.58

Standard errors in parentheses

Continued on next page

Table S6 – *Continued from previous page*

† significant at $p < .10$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

As final check, Table S6 predicts individual Justice votes using vocal pitch and the thirty best predictors form the {Marshall}+ algorithm. We used the thirty most pronounced predictors to make the results more interpretable. The authors (Katz, Bommarito and Blackman 2014) estimate their parameters using regression trees. This makes including all of their variables difficult. Moreover, the authors indicate most of their variables do not dramatically influence the model. In fact, the weights associated with the variables not found in Table S6 are essentially zero. As you can see, our variable ("Pitch Difference") is still highly significant ($p < 0.001$) even when these thirty variables are included as controls. A specification with all 95 variables yields the same result.

As explained in the paper, these results are not meant to suggest that vocal pitch is the only variable scholars should use when assessing emotional expression on the Supreme Court. The {Marshall}+ algorithm, text-based measures, and the "petitioner always wins" rule can (and should) be used to predict Justice votes. Our results simply suggest vocal pitch should also be included in this list, because it seems to carry considerable statistical and substantive significance.

Questions and Non-Questions

Below we report several robustness checks. Table S7 shows several different specifications of the models outlined in our paper. Beginning with Panel A, we show when one only uses questions the results are identical, regardless of whether one uses questions directed only at the petitioner (see Column 1) or questions directed only at the respondent (see Column 2). In fact, Column 3 shows when both are included in the same model you find both variables are highly significant ($p < 0.001$). In this model, the coefficients suggest when Justices speak at a higher vocal pitch towards the petitioner the petitioner is more likely to lose, whereas the same is true for questions directed towards the respondent.

Panel B replicates the petitioner results, but does not replicate the results for the respondent. In these models, we only include statements. These results emphasize the importance of questions when it comes to emotional activation. When Justices are asking questions they are actively trying to acquire new information. Mood congruent processing becomes much more likely during this process, suggesting Justices are more actively sorting information using their priors. The lack of a statistically significant relationship for non-questions further underlines this argument, even though consistent predictions are still found for questions directed towards the petitioner.

Table S7: Considering Petitioner Pitch and Respondent Pitch Separately

(a) Questions

	Petitioner Only	Respondent Only	Petitioner and Respondent
	(1)	(2)	(3)
Constant	0.13***	0.24***	0.16***
	(0.02)	(0.03)	(0.03)
Pet. Pitch	−0.16***		−0.28***
	(0.03)		(0.04)
Res. Pitch		0.11***	0.25***
		(0.03)	(0.04)
N	7,036	6,554	5,098
Log Likelihood	−4,845.04	−4,492.75	−3,484.95
AIC	9,694.08	8,989.50	6,975.90

(b) Non-Questions

	Petitioner Only	Respondent Only	Petitioner and Respondent
	(1)	(2)	(3)
Constant	0.11***	0.26***	0.20***
	(0.02)	(0.02)	(0.03)
Pet. Pitch	−0.12***		−0.15***
	(0.03)		(0.03)
Res. Pitch		0.02	0.04
		(0.02)	(0.03)
N	7,324	6,544	5,003
Log Likelihood	−5,055.31	−4,481.05	−3,430.79
AIC	10,114.62	8,966.09	6,867.57

(c) Both

	Petitioner Only	Respondent Only	Petitioner and Respondent
	(1)	(2)	(3)
Constant	0.12***	0.25***	0.18***
	(0.02)	(0.02)	(0.02)
Pet. Pitch	−0.14***		−0.19***
	(0.02)		(0.03)
Res. Pitch		0.05***	0.11***
		(0.02)	(0.02)
N	14,360	13,098	10,101
Log Likelihood	−9,900.95	−8,977.17	−6,926.01
AIC	19,805.91	17,958.33	13,858.01

(d) Both with Controls

	Petitioner Only	Respondent Only	Petitioner and Respondent
	(1)	(2)	(3)
Constant	0.23***	0.30***	0.24***
	(0.02)	(0.02)	(0.03)
Pet. Pitch	−0.13***		−0.18***
	(0.02)		(0.03)
Res. Pitch		0.05***	0.11***
		(0.02)	(0.02)
Lib. Petitioner	−0.25***	−0.14***	−0.16***
	(0.03)	(0.04)	(0.04)
Ideology	0.14***	0.18***	0.16***
	(0.01)	(0.01)	(0.02)
Lib. Petitioner × Ideology	−0.17***	−0.29***	−0.23***
	(0.02)	(0.02)	(0.02)
N	14,360	13,098	10,101
Log Likelihood	−9,804.93	−8,852.79	−6,854.41
AIC	19,619.86	17,715.58	13,720.82

Note: Outcome is whether the Justice voted in favor of petitioner. Unit of analysis is each Justice's vote. Models in panel A includes only statements with question marks ("Question" = 1). Models in panel B include only statements without question marks ("Question" = 0). Models in panels C and D include both, with models in the latter controlling for ideology. The average vocal pitch directed towards the petitioner ("Pet. Pitch") and the average vocal pitch directed towards the respondent ("Res. Pitch") are included as predictors. The column names indicate which combinations are used. Levels of significance are reported as follows: *p < .1; **p < .05; ***p < .01. Standard errors are reported in parentheses. All models include Justice fixed effects.

From this point, both Panels C and D demonstrate when questions and non-questions are considered simultaneously our results are extremely robust. In fact, every model yields highly significant results regardless of the specification. Beginning with Panel C, one finds essentially the same results as Panel A. When Justices raise their vocal pitch towards petitioners, the petitioners are more likely to lose (see Column 1). The same can be said for respondents (see Column 2), and each of these relationships hold when both are included in the same model (see Column 3). Panel D shows these results also hold even when Justice ideology is included as a control.

References

Black, Ryan C., Sarah A. Treul, Timothy Johnson and Jerry Goldman. 2011. "Emotions, Oral Arguments, and Supreme Court Decision Making." *The Journal of Politics* 73(2):572–581.

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.

Collins, Paul. 2004. "Friends of the Court: Examining the Influence of Amicus Curiae Participation in U.S. Supreme Court Litigation." *Law and Society Review* 38(4):807–832.

Collins, Paul. 2008. *Friends of the Supreme Court: Interest Groups and Judicial Decision Making.* Oxford University Press.

Epstein, Lee, Andrew D Martin, Kevin M Quinn and Jeffrey A Segal. 2007. "Ideological Drift Among Supreme Court Justices: Who, When, and How Important." *Northwestern University Law Review* 101(4):1483.

Farnsworth, Ward. 2007. "The Use and Limits of Martin-Quinn Scores to Assess Supreme Court Justices, with Special Attention to the Problem of Ideological Drift." *Northwestern University Law Review* 101(4):1891.

Katz, Daniel Martin, Michael James Bommarito and Josh Blackman. 2014. "Predicting the Behavior of the Supreme Court of the United States: A General Approach." *Available at SSRN 2463244* .

Martin, Andrew D. and Kevin M. Quinn. 2002. "Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953-1999." *Political Analysis* 10(2):134–153.

Martin, Andrew D., Kevin M. Quinn, Theodore W. Ruger and Pauline T. Kim. 2004. "Competing Approaches to Predicting Supreme Court Decision Making." *Perspectives on Politics* 2(4):761–767.

Rohde, David W. and Harold J. Spaeth. 1976. *Supreme Court Decision Making*. San Francisco, CA: W. H. Freeman.

Schubert, Glendon. 1965. *The Judicial Mind: The Attitudes and Ideologies of Supreme Court Justices, 1946-1963*. Evanston, IL: Northwestern University Press.

Schubert, Glendon. 1974. *The Judicial Mind Revisited: Psychometric Analysis of Supreme Court Ideology*. London, UK: Oxford University Press.

Segal, Jeffery A. and Albert D. Cover. 1989. "Ideological Values and the Votes of U.S. Supreme Court Justices." *American Political Science Review* 83:557–565.

Segal, Jeffrey A. 1988. "Amicus Curiae Briefs by the Solicitor General During the Warren and Burger Courts: A Research Note." *The Western Political Quarterly* 41(1):135–144.

Songer, Donald R. and Reginald S. Sheehan. 1993. "Interest Group Success in the Courts: Amicus Participation in the Supreme Court." *Political Research Quarterly* 46(2):339–354.

Spaeth, Harold J., Lee Epstein, Andrew D. Martin, Jeffrey A. Segal, Theodore J. Ruger and Sara C. Benesh. 2015. "Supreme Court Database, Version 2015, Release 01." `http://www.http://supremecourtdatabase.org` (Accessed on August 23, 2015).