PA Emotional Arousal Predicts Voting on the U.S. Supreme Court

Bryce J. Dietrich[1], Ryan D. Enos[2] and Maya Sen[3]

[1] Department of Political Science, University of Iowa, 341 Schaeffer Hall, Iowa City, IA 52242, USA.
Email: bryce-dietrich@uiowa.edu, URL: http://www.brycejdietrich.com
[2] Department of Government, Harvard University, 1737 Cambridge St., Cambridge, MA 02138, USA.
Email: renos@gov.harvard.edu, URL: http://ryandenos.com
[3] John F. Kennedy School of Government, Harvard University, 79 John F. Kennedy Street, Cambridge, MA 02138, USA.
Email: maya_sen@hks.harvard.edu, URL: http://scholar.harvard.edu/msen

Abstract

Do judges telegraph their preferences during oral arguments? Using the U.S. Supreme Court as our example, we demonstrate that Justices implicitly reveal their leanings during oral arguments, even before arguments and deliberations have concluded. Specifically, we extract the emotional content of over 3,000 hours of audio recordings spanning 30 years of oral arguments before the Court. We then use the level of emotional arousal, as measured by vocal pitch, in each of the Justices' voices during these arguments to accurately predict many of their eventual votes on these cases. Our approach yields predictions that are statistically and practically significant and robust to including a range of controls; in turn, this suggests that subconscious vocal inflections carry information that legal, political, and textual information do not.

Keywords: audio, supreme court, prediction

1 Introduction

In February of 2017, the Trump Administration issued an executive order that banned the entry of people from seven majority-Muslim countries into the U.S. and unleashed chaos in airports all over the world. Legal challenges were swift and, within two weeks, the 9th Circuit Court of Appeals had scheduled oral arguments concerning the order. Public interest in how the court would rule was significant. More than 130,000 people listened to the arguments, and hundreds of experts weighed in on how the judges would vote. Many of these predictions relied on the three judges' emotional reactions and vocal expressions during oral argument. For example, the *New York Times* provided a live analysis of the judges' reactions, assessing whether they were "pretty skeptical" or "friendly."[1] Ultimately, the three-judge panel ruled unanimously against the travel ban, but it was not without days of uncertainty for those affected.

In this paper, we do what observers of the 9th Circuit oral argument were attempting by asking whether we can systematically detect how judges will vote based on emotional responses at oral argument. We address this by examining the U.S. Supreme Court, which has decades of audio data. Specifically, we explore audio from nearly 3,000 hours of oral arguments from the last 30 years and find that *vocal pitch alone* is strongly predictive of Supreme Court Justices' votes. The results are robust to the inclusion of other factors and predict outcomes at least as well as more complex models accounting for substantive features of the actors and cases, suggesting that vocal pitch predicts decisions in ways that characteristics like ideology or legal issue areas do not. In results presented in the Supplemental Information, moreover, we extend our findings to the 9th Circuit's travel ban argument, suggesting that our approach has good external validity.

1 https://www.nytimes.com/interactive/2017/02/07/us/ninth-circuit-oral-arguments-trump-immigration.html.

As we discuss below, our findings are consistent with several causal narratives. For example, it could be the case that Justices actively rely on their emotions in reaching important decisions. It may also be that judges experience aroused responses during oral arguments because they receive information conflicting with previously made decisions. Although we cannot disaggregate these explanations, our results clearly show that nonsubstantive and implicit signals, even among elite actors such as federal judges and Supreme Court Justices, can provide additional meaningful information on their attitudes beyond what can be found in their textual pronouncements.

2 Description of Supreme Court Oral Arguments Audio Data, Emotional Arousal, and Vocal Pitch

We are not the first to suggest that emotion plays an important role in judicial oral arguments (e.g., Shullman 2004; Johnson *et al.* 2009). For example, Black *et al.* (2011) argue that the "tenor" of oral arguments can be used as a barometer of how Justices will rule in a given case. While others have utilized the number of questions directed toward each side (Roberts, Jr. 2005; Epstein, Landes, and Posner 2010), Black *et al.* (2011) captured the emotions expressed during oral arguments using the number of "pleasant" and "unpleasant" words. They find that the more unpleasant words Justices use toward an attorney, the less likely that attorney will prevail in the case. Even though we acknowledge the importance of what Justices say, we suggest *how* they say those words may be of equal, if not greater, predictive importance.

Why might vocal pitch predict the behavior of even elite actors like Supreme Court Justices and why might the emotion signaled by vocal pitch be more important than the substantive content? First, changes in vocal inflections, like pitch, often occur unbeknownst to the speaker (Ekman *et al.* 1991). For the Justices, emotional arousal may be more likely when interacting with someone with whom they disagree. When this occurs, the heart begins to race, palms begin to sweat, and all muscles, including the vocal cords, tighten (Posner, Russell, and Peterson 2005). This is the primary reason why "the most consistent association 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). For example, Laukka, Juslin, and Bresin (2005) asked actors to portray "weak" and "strong" versions of a variety of emotions that were later judged by amateur and expert judges on the degree to which the actors were displaying an "activated" or "intense" emotional state. Not only was vocal pitch correlated with both activation and intensity, but the "patterns of vocal cues for activation and emotion intensity showed numerous similarities," suggesting both may be capturing the same "physiological reaction" (648).[2]

Moreover, due to the automatic nature of this response, a speaker's vocal pitch will often provide insights into a speaker's level of activation beyond their conscious communication. Indeed, "several studies have shown, that like the body, 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). (Additional discussion of this literature can be found in the Supplemental Information.) For these reasons, verbal and nonverbal behavior should be thought of in terms of a "leakage hierarchy" with "verbal content" 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 suggests that Justices may subconsciously indicate their ultimate preferences toward a case by raising their vocal pitch toward either the petitioner or respondent. This is the case regardless of whether the

---

2 In a similar study, Bänziger and Scherer (2005) found actors portraying emotions with "high" activation not only spoke with a higher average vocal pitch (mean F_0), but they also found little variation in the F_0 contours, suggesting that "simple summaries of F_0 contours—such as F_0 mean or F_0 range—were sufficient to account for the most important variations observed between categories" (265).

Justice formulates her response contemporaneously or whether she is reacting on the basis of predispositions about the case.

Second, some Justices, like Antonin Scalia, may be more willing to express emotion as compared to others. Such differences are problematic for text-based measures since they only capture emotion that is verbalized. According to Russell (2003), this ignores a range of emotions which occur prior to conscious awareness. As analogy, consider felt body temperature. Even though our body's temperature changes all the time, we do not always identify those changes as being *hot* or *cold*. For some, a small decrease in temperature may be enough to say, "I am cold!" For others, that same decrease may not even be recognized. Emotional expression on the Supreme Court functions in a similar way—some is verbal, some is nonverbal. For some Justices, an attorney's error may be egregious enough to warrant calling it "idiotic," while for others that same error may not even raise an eyebrow, verbal or otherwise. Text-based measures are extraordinarily useful when one is interested in understanding the former, but struggle with latter.[3]

To achieve this end, we collected audio recordings from oral arguments in 1,773 cases, beginning in 1982 and ending in 2014.[4] Using the timestamps provided by the Oyez Project,[5] we further parsed these cases into discrete segments of audio uttered by (1) the Justices themselves, (2) the lawyer/s representing the petitioner, and (3) the lawyer/s representing the respondent.[6] Lawyers spoke for 2,137 hours. Justices spoke for 502 hours. For the Justices, this represented 146,335 discrete utterances. Additional descriptive statistics are provided in the Supplemental Information, Tables S1–S2.

3 Results: How Emotion Arousal Predicts Supreme Court Justices' Voting

We expect that a Justice who is more emotionally activated when speaking toward an attorney will be more likely to vote against that attorney. If this is correct, a higher vocal pitch will predict a stronger negative response. To analyze this, we code whether a Justice votes in favor of the petitioner, a 1 or 0 variable.[7] We construct a measure of "Pitch Difference" by subtracting vocal pitch in questions directed toward petitioners from vocal pitch in questions directed toward respondents. (Vocal pitch was measured using *Praat*, a speech synthesis program that estimates the fundamental frequency by dividing the autocorrelation of a windowed signal by the autocorrelation of the window itself. To estimate the fundamental frequency we only use voiced speech. More details can be found in the Supplemental Information.) For each Justice, we converted vocal pitch to standard deviations above and below his or her average vocal pitch, which accounts for systematic differences between Justices (for example between male and female Justices) as well as any measurement error associated with extracting the fundamental frequency.[8]

---

3 Note that some nonverbal responses could still be consciously made—for example, a Justice explicitly raising the tone of her voice to talk over another Justice. Our approach actually gauges the predictive power of both conscious and subconscious nonverbal expression. However, as we note in the SI, we have reason to think that much (though not certainly all) variation in vocal pitch operates in a subconscious level and is uncontrollable.

4 We have made the data and the code needed to replicate our study publicly available in Dataverse. Please see Dietrich, Enos, and Sen (2017) for more details.

5 https://www.oyez.org. More information on the audio data acquisition is in the Supplemental Information.

6 The "petitioner" is the party bringing the case to the Court; the "respondent" is the party responding to the petitioner's claim. Both parties are represented by separate counsel, often experienced attorneys from the Supreme Court bar.

7 Since the outcome is 1 (Justice votes for the petitioner) or 0 (Justice votes against the petitioner), we use a multilevel logistic regression, which we implemented here via the `lme4` package in R statistical software language.

8 We explore these issues in the "Estimating the Fundamental Frequency" section of the Supplemental Information. Since male vocal cords are typically longer (17.5–25 mm) than female vocal cords (12.5–17.5 mm), men tend to talk at a lower vocal pitch—making standardization essential for comparing male and female Justices. Such standardization also helps account for errors that may occur when estimating the fundamental frequency, although we show this is unlikely to be consequential when estimating mean vocal pitch.

The main results are presented in Table 1, Model 2. All models are multilevel logistic regressions with random intercepts for each Justice.[9] These results show *that the higher emotional arousal or excitement directed at an attorney compared to his or her opponent, the less likely that attorney is to win the Justice's vote* ($p < 0.001$).[10] From Model 1, when the vocal pitch of questions directed to both sides is the same, the predicted probability of a Justice voting for the petitioner is 0.55. However, the probability of a Justice voting for the petitioner drops by 7 percentage points if the difference between the vocal pitch directed to the petitioner is one standard deviation higher than the vocal pitch directed at the respondent. The overall prediction rate is also reported.[11] Here, we are able to predict 57.50 percent of Justices' votes accurately (see Table 1) and 66.55 percent of overall case outcomes accurately (see Table S4) using only pitch difference, suggesting that vocal pitch predicts not only how individual Justices vote but also the eventual disposition of the case.[12]

Models 2, 3, and 4 include the controls used by Black *et al.* (2011), as well as the differences in the use of "pleasant" and "unpleasant" words as defined by the Dictionary of Affect in Language (DAL), the Harvard IV dictionary (also known as the General Inquirer), and the Linguistic Inquiry and Word Count (LIWC) dictionary, respectively. Since the Harvard IV dictionary is publicly available, we provide the words used for Model 3 in the Supplemental Information. For Models 2 and 4, we provide some examples of "positive" and "negative" words. Unlike Black *et al.* (2011), we use the Martin–Quinn scores estimated in the previous term, as Martin–Quinn scores are dynamically estimated within each term using Justices' votes, which would introduce endogeneity if not lagged. These are continuous measures from liberal (−1.0) to conservative (+1.0) and vary from Justice to Justice and from term to term.

After accounting for vocal pitch, including these other variables only increases the predictive power of the model by around seven percentage points, suggesting that vocal pitch has unique predictive value. To further assess the substantive importance of vocal pitch, we compared the performance of vocal pitch (and of only vocal pitch) to a widely known algorithm developed by Katz, Bommarito, and Blackman (2014), known as {Marshall}+. This algorithm uses 95 variables to predict Supreme Court Justices' voting and is known as one of the most predictive algorithms currently available.[13] As explained in the Supplemental Information, we restricted our analysis to the period from 1998 to 2012. In total, {Marshall}+ successfully predicts 64.76 percent of cases correctly, which is 1.79 percentage points *lower* than our prediction rate of 66.55 percent of cases. A simple χ^2 test reveals the models are similarly predictive ($\chi^2 = 0.52$, $df = 1$, $p > 0.05$), suggesting we are able to equal the predictive power of a model that uses 95 predictors using only one—vocal pitch.

Model 1 also outperforms traditional petitioner-based models in which a "plaintiff always wins" rule is applied. Although seemingly simple, such a rule is actually fairly sophisticated and takes into account a lot of what scholars know about strategic planning, the rule of four, and principles of precedent. Using only vocal pitch, Model 1 significantly ($p \leq 0.001$) outperforms this baseline by 2.58 percentage points. Even though Models 2, 3, and 4 all perform better, when one only uses the number of "positive" and "negative" words, the prediction rates are substantially worse. For example, when the only predictors are the percent more positive and negative words directed

---

9 Due to space limitations, we do not describe all the control variables in the main text. Full descriptions and additional model specifications can be found in the Supplemental Information.

10 Since we cannot assume that Justices' votes within cases are independent, we include a randomly varying intercept for each Justice. This additional parameter also helps account for other within-Justice differences.

11 We used a 0.50 threshold for these calculations, meaning when the model returned a predicted probability greater than 0.50, we predicted the Justice would vote for the petitioner.

12 In the Supplemental Information, we also report the true positive, true negative, false positive, and false negative rates—comparing each to a variety of previous studies. These statistics are included as part of a broader discussion of how best to assess predictive performance.

13 These 95 variables include case information, ideological information, Supreme Court trends, Justice background characteristics, etc.). See http://lexpredict.com/portfolio/predicting-the-supreme-court.

Table 1. Does vocal pitch predict votes in favor of the petitioner?

	Intercept only	No controls		Harvard	
			DAL	IV	LIWC
	(1)	(2)	(3)	(4)	(5)
Fixed effects					
Constant	0.202 (0.056)	0.178 (0.055)	−0.025 (0.160)	−0.027 (0.160)	−0.026 (0.160)
Pitch difference		−0.266 (0.036)	−0.214 (0.038)	−0.215 (0.038)	−0.214 (0.038)
% More unpleasant words directed at petitioner			−1.971 (1.471)	0.071 (0.846)	−2.129 (1.308)
% More pleasant words directed at petitioner			−1.647 (1.086)	0.272 (0.685)	−1.673 (1.046)
# More questions directed at petitioner			−0.057 (0.008)	−0.057 (0.008)	−0.057 (0.008)
Political ideology$_{t-1}$			0.158 (0.033)	0.158 (0.033)	0.158 (0.033)
LC decision was conservative			0.011 (0.073)	0.012 (0.073)	0.012 (0.073)
Political ideology$_{t-1}$ × LC decision was conservative			−0.263 (0.034)	−0.264 (0.034)	−0.263 (0.034)
Solicitor general as amicus supporting petitioner			0.540 (0.079)	0.543 (0.079)	0.540 (0.079)
Solicitor general as amicus supporting respondent			−0.672 (0.104)	−0.666 (0.104)	−0.673 (0.104)
# of amicus briefs supporting petitioner			0.039 (0.008)	0.039 (0.008)	0.039 (0.008)
# of amicus briefs supporting respondent			−0.058 (0.007)	−0.058 (0.007)	−0.058 (0.007)
Petitioner's level of resources			0.045 (0.014)	0.046 (0.014)	0.045 (0.014)
Respondent's level of resources			−0.003 (0.014)	−0.003 (0.014)	−0.003 (0.014)
Random effects					
Intercept	0.03 (0.16)	0.03 (0.16)	0.02 (0.13)	0.02 (0.12)	0.02 (0.12)
N_1	5,209	5,209	4,977	4,977	4,977
N_2	18	18	18	18	18
log L	−3,581.134	−3,551.721	−3,201.611	−3,203.521	−3,201.128
AIC	7,166.268	7,109.441	6,433.222	6,437.041	6,432.256
Percent correctly predicted	54.92	57.50	63.91	63.49	63.67

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 toward the petitioner ("Petitioner Pitch") minus the average vocal pitch in questions directed toward 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). Please refer to the Supplemental Information for more details about each dictionary, the controls, and our cross-validation approach.

at the petitioner, the model successfully predicts 0.81 and 0.72 percentage points better than the "plaintiff always wins" model, depending on whether one uses the DAL or LIWC dictionaries, respectively. The Harvard IV dictionary actually predicts 8.25 percentage points *less* than this baseline. As we show in the Supplemental Information, regardless of the text-based measure one uses, vocal pitch does substantially better at predicting both cases and votes. These results are not meant to suggest that vocal pitch is the *only* variable that should be incorporated into models of oral arguments. Nor do our results suggest text-based measures have no place in the study of emotional expression on the Supreme Court. Rather, our results demonstrate vocal pitch should be one of many variables that should be taken into consideration when assessing oral arguments.

4 Discussion

For scholars interested in predicting Justice votes weeks, if not months, before the Court's ruling is released, oral arguments "provide a barometer of how justices will rule in a given case" (Black *et al.* 2011, 574). While we are not the first to suggest emotional expressions are an important part of such prognostications, the vast majority of these studies have only considered text-based measures (for important exceptions, see Schubert *et al.* 1992; Knox and Lucas 2017). We show vocal pitch on its own is about as predictive of Justices' votes and overall case outcomes as models that use all publicly available quantitative legal and nonlegal information, including additional textual information related to emotion. These comparisons are not meant to suggest that vocal pitch is the *only* variable scholars should use when assessing emotional expression on the Supreme Court. We argue the {Marshall}+ algorithm, text-based measures, and the "petitioner always wins" rule can (and should) be used to predict Justice votes. However, nonverbal signals, including changes in vocal pitch, also carry considerable weight. Justices choose their words carefully, but have far less control over how those words are spoken—and these subconscious vocal cues, our findings show, carry important information about eventual rulings.

Supplementary material

For supplementary material accompanying this paper, please visit
https://doi.org/10.1017/pan.2018.47.

References

Bänziger, Tanja, and Klaus R. Scherer. 2005. The role of intonation in emotional expressions. *Speech Communication* 46(3):252–267.

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.

Dietrich, Bryce J., Ryan D. Enos, and Maya Sen. 2017. Emotional arousal predicts voting on the U.S. supreme court. https://doi.org/10.7910/DVN/JFU71R, Harvard Dataverse, V1.

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.

Epstein, Lee, William M. Landes, and Richard A. Posner. 2010. Inferring the winning party in the supreme court from the pattern of questioning at oral argument. *The Journal of Legal Studies* 39(2):433–467.

Johnson, Timothy R., Ryan C. Black, Jerry Goldman, and Sarah A. Treul. 2009. Inquiring minds want to know: Do justices tip their hands with questions at oral argument in the U.S. supreme court? *Washington University Journal of Law & Policy* 29:241–261.

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: https://ssrn.com/abstract=2463244.

Knox, Dean, and Christopher Lucas. 2017. A general approach to classifying mode of speech: The speaker-affect model for audio data. Unpublished manuscript.

Laukka, Petri, Patrik Juslin, and Roberto Bresin. 2005. A dimensional approach to vocal expression of emotion. *Cognition & Emotion* 19(5):633–653.

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

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(3):715–734.

Roberts, John G. Jr. 2005. Oral advocacy and re-emergence of a Supreme Court bar. *Journal of Supreme Court History* 30:68–81.

Russell, James A. 2003. Core affect and the psychological construction of emotion. *Psychological Review* 110:145–172.

Schubert, James N., Steven A. Peterson, Glendon Schubert, and Stephen Wasby. 1992. Observing supreme court oral argument: A biosocial approach. *Politics and the Life Sciences* 11(1):35–51.

Shullman, Sarah Levien. 2004. The illusion of Devil's advocacy: How the justices of the supreme court foreshadow their decisions during oral argument. *Journal of Appellate Practice and Process* 6:271–293.

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 Erlbaumb, pp. 129–147.