

Attorney Voice and the U.S. Supreme Court

Daniel L. Chen^{*}, Yosh Halberstam[†],
Manoj Kumar[‡] and Alan Yu[§]

Introduction

The natural audio presentation of natural language has many sources of various beyond simply the choice of words. Characteristics of a speech act such as pitch, diction, and intonation may be significant even though they do not affect the semantic content of what has been spoken. There is a significant body of scholarship that examines this type of speech variation, e.g., in mate selection, leader selection, housing choices, consumer purchases, and even stock market outcomes (Nass and Lee 2001; Klofstad, Anderson, and Peters 2012; Purnell, Idsardi, and Baugh 1999; Scherer 1979; Tigue et al. 2012; Mayew and Venkatachalam 2012), but there is relatively little quantitative empirical evidence that speech variation beyond lexical choices matters for real-world behavior. Speech variation from identical utterances of ‘Hello’ affect personality ratings

^{*}Toulouse School of Economics, Institute for Advanced Study in Toulouse, University of Toulouse Capitole, Toulouse, France

[†]University of Toronto, Economics Department

[‡]NYU Center for Data Science

[§]University of Chicago, Linguistics Department

(McAleer, Todorov, and Belin 2014), but linking these ratings to downstream behavior is challenging. Nevertheless, oral advocacy classes are taught at law schools and skilled oral advocacy is a highly sought after professional trait (Korn 2004).

In this paper, we take up the question of the practical relevance of speech variation by examining whether specific vocal cues in the first three seconds of speech are predictive in high-stakes policy-making settings such as the U.S. Supreme Court.

There are many reasons to think that vocal first impressions should not matter very much. From the perspective of a purely rational judicial decision maker, only the information content of a speech act should count (Posner 1973). Unless vocal characteristics carry useful information, they should be ignored. For example, under the attitudinal model, judicial decision making is understood as largely political, with outcomes determined by judicial attributes (Cameron 1993). Something as seemingly insignificant as vocal characteristics should not be enough to overwhelm a judge's ideological dispositions. Alternatively, under a legal model, judges would focus on the legal content of the arguments presented by litigants (Kornhauser 1999). Vocal style would again seem to be irrelevant.

More broadly, even if vocal style did influence judge's decisions, competitive pressures should work to eliminate low cost arbitrage opportunities (Becker 2010). Specifically, if it became known that judges prefer a certain vocal style, and it was relatively low cost to adopt that vocal style, then all lawyers would do so or risk losing clients. By analogy, it is relatively low cost for attorneys to dress according to the norms and standards of the profession by donning business suits in courts. An attorney who simply refused to conform to the norm by wearing jeans to court would risk unnecessarily biasing judges in ways that were adverse to his or her clients (and perhaps even run the risk of a malpractice lawsuit). For this reason, lawyers wear suits to court, even if that attire is not to every one of their personal tastes. At the very least, a lawyer who refuses to

wear a suit would likely have to discount the cost of his or her services to reflect the lower value for clients, to the point where differences between attorney attire would be “priced into” the marketplace, and therefore, would no longer be observed.

However, if firms or clients have a preference for certain advocates beyond their performance in court, correlations between malleable characteristics and outcomes could persist. For example, if courts tended to be biased against lawyers who adopt certain behaviors, but clients insisted on hiring on an equal opportunity basis, then those lawyers might under-perform compared to others at the same pay scale. Alternatively, judges may be unbiased, but law firms and clients could be biased in their hiring practices, leading to the under-performance of lawyers in the favored class. Indeed, legal theorists have suggested that discrimination, once aimed at entire groups based on “immutable characteristics,” now aims at subsets that refuse to *cover*—i.e., refuse to assimilate their behavior to dominant norms (Yoshino 2006; Goffman 1963). Yoshino (2006) argues that when courts allow employment that is contingent on covering, it legitimizes second-class citizenship for the subordinated group. Subordination would conflict with values expressed in the U.S. Constitution (Balkin 2011).

The question of covering-based discrimination is just beginning to attract attention from empirical scholars (Bertrand and Duflo 2016; Neumark 2016). When it comes to how one speaks, minorities’ choice of diction have been found to be associated with long-run labor market outcomes (Grogger 2011). Female lawyers routinely pay coaches to sound more masculine (Starecheski 2014), an indication at the very least of a perception of discrimination. Men with non-masculine voices have also been found to be disadvantaged in the labor market (Case 1995).

The following analysis examines whether vocal characteristics have predictive power in a setting in which there are very powerful market forces that would seem well-poised to arbitrage away any low-cost competitive advantage: oral arguments before the U.S. Supreme Court. This

work reports and builds on prior analyses (Chen, Halberstam, and Yu 2016a, 2016b) and is based on the data of 1,901 U.S. Supreme Court oral arguments between 1998 and 2012. Specifically, we will examine whether voice-based snap judgments based on lawyers' identical introductory sentences, "Mr. Chief Justice, (and) may it please the Court?" predict court outcomes.

Data

The data come from Chen, Halberstam, and Yu (2016a). Oral arguments at the Supreme Court have been recorded since the installation of a recording system in October 1955. The recordings and the associated transcripts are made available to the public in electronically downloadable format by the Oyez Project (<http://www.oyez.org/>), a multimedia archive at the Chicago-Kent College of Law devoted to the Supreme Court and its work. The audio archive contains more than 110 million words in more than 9000 hours of audio synchronized, based on the court transcripts. Oral arguments are, with rare exceptions, the first occasion in the processing of a case in which the Court meets in person with the litigants' counsel to consider the issues. Usually, counsel representing the competing parties of a case each have 30 minutes to present their side to the Justices. The Justices may interrupt these presentations with comments and questions, leading to interactions involving the Justices, the lawyers and, in some cases, the amici curiae. All audio clips involve the lawyers' opening statements. The first handful of words in those statements are identical: "Mr. Chief Justice, (and) May It Please the Court".

The labeled sample comprise almost 2,000 Supreme Court advocate audio clips for 15 years from 1998–2012 with ratings for confidence, masculinity, trust, intelligence, attractiveness and aggressiveness. Each audio clip was rated by approximately 20 Mechanical Turk workers, and a total of 20,888 ratings are available in this database. This data serves as a training set for 14,932 unrated audio clips of Supreme Court advocates from 1946–1997 and 2013–2014, spanning roughly 70 years.

The raters were asked to use headphones and to rate on a Likert scale from 1 (low) to 7 (high) the characteristics of masculinity, attractiveness, confidence, intelligence, trustworthiness, and aggressiveness. These six traits were selected based on previous research on listeners' perceptual evaluations of linguistic variables (Eckert 2008; Campbell-Kibler 2010; McAleer, Todorov, and Belin 2014). They are also similar to the ones used in Todorov et al. (2005), which presented subjects with pictures of electoral candidates' faces and asked them to rate their perceived attributes. That study found that perceptions of competence predicted election outcomes. Male and female lawyers were rated in separate blocks, such that participants either rated male advocates or female advocates but not both, so raters would not be comparing females and males on the degree of masculinity. Female lawyers were rated in terms of femininity instead of masculinity.

The elicitation algorithm randomized the order of the questions and whether "masculine" or "feminine" occupied the left or right portion of the scale (i.e. the *polarity* of the scale). The order and polarity of questions were held fixed for any particular rater to minimize cognitive fatigue. For additional nudges across experimental designs and to ensure attention by the rater, listening attention checks were employed. If raters failed they would be dropped from the sample. There were six alertness trials, three with beeps and three without. The beep comes at the beginning of the lawyer's voice. For these questions subjects were asked if they heard a beep, but not to rate the lawyer's voice.

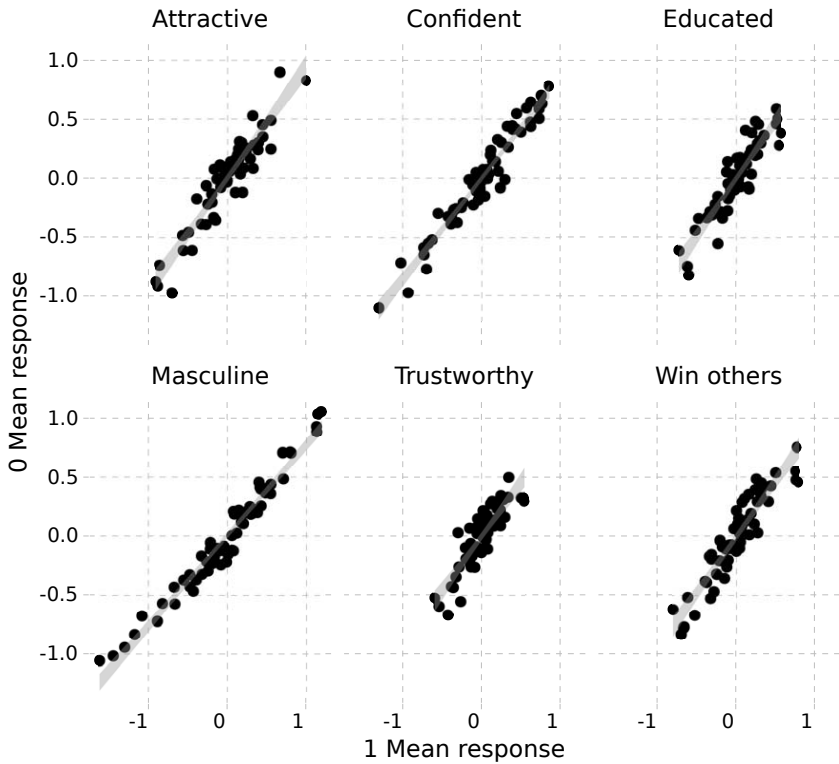
Raters were also asked to rate the quality of the recording. While there is no time limit on how long a subject can spend on each trial, they were given a minimum of 5 seconds to respond; they were not allowed to proceed to the next trial until the 5 seconds was up (and all the questions completed) in order to ensure that subjects were given enough time to complete the ratings and to discourage them from speeding through the trials. No information regarding the identity of the lawyer or the nature of the case was given to the participants.

To control for the possibility of within-voice modeling by raters, instead of the basic design (in which the listener was presented with one voice sample and rates the sample on all scales), Chen, Halberstam, and Yu (2016b) also employed a design with only one question, randomly selected for each voice sample with only 60 clips and fewer subjects. Each voice clip was played aloud only once, in order to capture the respondents' first impressions and to avoid them over thinking their responses (Ballew and Todorov 2007). There was a high degree of correlation of individual perceptions across experimental designs and stimulus presentation methods (see Figure 1).

Figure 1 plots the mean rating for each of the advocates using the two approaches discussed above. The x-axis reflects the the mean ratings obtained from raters who were asked to rate each advocate on the full set of attributes, and the y-axis reflects the mean ratings obtained from raters who were randomly assigned to rate each advocate on only one attribute. The ratings are highly correlated across these experimental designs, suggesting that trait judgments obtained from listening to a voice are quite stable.

Additional data for predicting Supreme Court outcomes was drawn from Katz, Bommarito, and Blackman (2014) which sought to predict Supreme Court decisions using pre-trial characteristics (which collectively received roughly 25% of the importance weight) as well as court and judge historical trends specific to issue, parties, and lower courts (which collectively received roughly 75% of the importance weight). These features are divided into seven categories and presented in Table 1: a) Justice and Court Background Information (e.g., Justice year of birth), b) Case Information (e.g., Legal Issue), c) Overall Historic Supreme Court Trends (e.g., Ideological Direction), d) Lower Court Trends (e.g., Circuit Court Ideological Trend), e) Current Supreme Court Trends (e.g., Mean Agreement Level of Current Court), f) Individual Supreme Court Justice Trends (e.g., Mean Justice Ideological Direction), and g) Differences in Trends (e.g., Difference Between Justice and Circuit Court Directions).

Figure 1. *Correlation in Average Voice Perceptions across Experiments (Many vs. One Attribute)*



Notes: This figure plots the mean rating of sixty voice samples in the pilot, where the x-axis reflects mean ratings obtained from raters who were asked to rate each advocate on the full set of attributes, whereas the y-axis reflects the mean ratings obtained from raters who were randomly assigned to rate each advocate on only one attribute. Standard error is estimated with a linear model.

In Katz, Bommarito, and Blackman (2014) random forest—a weighted non-parametric model that forms weighted predictions based on nearest neighbors—was found to perform well.

1 Baseline Model and Performance Evaluation

Chen, Halberstam, and Yu (2016a) used the Katz, Bommarito, and Blackmun prediction of Justices' votes as a control in a linear regression. Chen, Halberstam, and Yu asked whether vocal characteristics had an explanatory effect above and beyond the predictors generated by the Katz, Bommarito, and Blackmun model, and if so, how much. That paper argued that the reason vocal characteristics performed well relative to the best prediction model is that, as noted in Katz, Bommarito, and Blackman (2014), that model performs best on cases on which the Court was in agreement (9-0) and performs worst on cases with high levels of disagreement among members of the Court (5-4). In fact, in close cases affirming the lower court, the model predicts the outcome with only 25% accuracy.¹ Chen, Halberstam, and Yu (2016a) showed that vocal characteristics are predictive of outcomes with the swing voter, which is where the random forest model may do poorly. In brief, the vast majority of judge votes are in easy cases, where extra-legal and extra-ideological factors may play a smaller role. In hard cases, where judges are closer to indifference (i.e. for close calls), human biases could tip the swing vote, whose importance is magnified when examining case outcomes.

The current paper extends the analysis to the available universe of Supreme Court oral arguments. Again, the best existing predictive model is used as a baseline. Features are then added to determine whether they increase accuracy. Chen, Halberstam, and Yu (2016a) showed that vocal characteristics are predictive of Supreme Court votes depending on the

1. Figure 6 in Katz, Bommarito, and Blackman (2014)

Table 1. Final Feature Weights (Katz, Bommarito, and Blackman 2014)

Justice and Court Background Information		Current Supreme Court Trends	
Justice [S]	0.00781	Mean Agreement Level of Current Court [FE]	0.00955
Justice Gender [FE]	0.00205	Std. Dev. of Agreement Level of Current Court [FE]	0.00936
Is Chief [FE]	0.00283	Mean Current Court Direction Circuit Origin [FE]	0.00789
Party President [FE]	0.00604	Std. Dev. Current Court Direction Circuit Origin [FE]	0.00850
Natural Court [S]	0.00764	Mean Current Court Direction Circuit Source [FE]	0.00945
Segal Cover Score [SC]	0.00971	Std. Dev. Current Court Direction Circuit Source [FE]	0.01021
Year of Birth [FE]	0.00793	Mean Current Court Direction Issue [FE]	0.01469
TOTAL	0.04403	Z-Score Current Court Direction Issue [FE]	0.00832
Case Information		Std. Dev. Current Court Direction Issue [FE]	0.01266
Admin Action [S]	0.00978	Mean Current Court Direction [FE]	0.00918
Case Origin [S]	0.00971	Std. Dev. Current Court Direction [FE]	0.00942
Case Origin Circuit [S]	0.00845	Mean Current Court Direction Petitioner [FE]	0.00863
Case Source [S]	0.00953	Std. Dev. Current Court Direction Petitioner [FE]	0.00894
Case Source Circuit [S]	0.01015	Mean Current Court Direction Respondent [FE]	0.00882
Law Type [S]	0.01370	Std. Dev. Current Court Direction Respondent [FE]	0.00888
Lower Court Disposition Direction [S]	0.01190	TOTAL	0.14456
Lower Court Disposition [S]	0.01125	Individual Supreme Court Justice Trends	
Lower Court Disagreement [S]	0.00706	Mean Justice Direction [FE]	0.01248
Issue [S]	0.01541	Mean Justice Direction 10 [FE]	0.01530
Issue Area [S]	0.01469	Mean Justice Direction Z Score [FE]	0.00826
Jurisdiction Manner [S]	0.00595	Mean Justice Direction Petitioner [FE]	0.00732
Month Argument [FE]	0.02014	Mean Justice Direction Petitioner 10 [FE]	0.01027
Month Decision [FE]	0.01349	Mean Justice Direction Respondent [FE]	0.00724
Petitioner [S]	0.01406	Mean Justice Direction Respondent 10 [FE]	0.01030
Petitioner Binned [FE]	0.01199	Mean Justice Direction for Circuit Origin [FE]	0.00792
Respondent [S]	0.01490	Mean Justice Direction for Circuit Origin 10 [FE]	0.00945
Respondent Binned [FE]	0.01179	Mean Justice Direction for Circuit Source [FE]	0.00891
Cert Reason [S]	0.01408	Mean Justice Direction for Circuit Source 10 [FE]	0.00970
TOTAL	0.22814	Mean Justice Direction by Issue [FE]	0.01881
Overall Historic Supreme Court Trends		Mean Justice Direction by Issue 10 [FE]	0.00950
Mean Court Direction [FE]	0.00988	Mean Justice Direction by Issue Z Score [FE]	0.00771
Mean Court Direction 10 [FE]	0.01997	TOTAL	0.14323
Mean Court Direction Issue [FE]	0.01546	Differences in Trends	
Mean Court Direction Issue 10 [FE]	0.00938	Difference Justice Court Direction [FE]	0.01210
Mean Court Direction Petitioner [FE]	0.00863	Abs. Difference Justice Court Direction [FE]	0.00929
Mean Court Direction Petitioner 10 [FE]	0.00904	Difference Justice Court Direction Issue [FE]	0.01167
Mean Court Direction Respondent [FE]	0.00875	Abs. Difference Justice Court Direction Issue [FE]	0.00968
Mean Court Direction Respondent 10 [FE]	0.00925	Z Score Difference Justice Court Direction Issue [FE]	0.01055
Mean Court Direction Circuit Origin [FE]	0.00791	Difference Justice Court Direction Petitioner [FE]	0.00705
Mean Court Direction Circuit Origin 10 [FE]	0.00864	Abs. Difference Justice Court Direction Petitioner [FE]	0.00708
Mean Court Direction Circuit Source [FE]	0.00951	Difference Justice Court Direction Respondent [FE]	0.00690
Mean Court Direction Circuit Source 10 [FE]	0.01017	Abs. Difference Justice Court Direction Respondent [FE]	0.00699
TOTAL	0.12663	Z Score Justice Court Direction Difference [FE]	0.01280
Lower Court Trends		Justice Lower Court Direction Difference [FE]	0.01922
Mean Lower Court Direction Circuit Source [FE]	0.00962	Justice Lower Court Direction Abs. Difference [FE]	0.02494
Mean Lower Court Direction Circuit Source 10 [FE]	0.01017	Justice Lower Court Direction Z Score [FE]	0.01126
Mean Lower Court Direction Issue [FE]	0.01334	Z Score Justice Lower Court Direction Difference [FE]	0.00992
Mean Lower Court Direction Issue 10 [FE]	0.00933	Agreement of Justice with Majority [FE]	0.00866
Mean Lower Court Direction Petitioner [FE]	0.00949	Agreement of Justice with Majority 10 [FE]	0.01483
Mean Lower Court Direction Petitioner 10 [FE]	0.00874	Difference Court and Lower Ct Direction [FE]	0.01522
Mean Lower Court Direction Respondent [FE]	0.00973	Abs. Difference Court and Lower Ct Direction [FE]	0.01199
Mean Lower Court Direction Respondent 10 [FE]	0.00900	Z-Score Difference Court and Lower Ct Direction [FE]	0.01217
TOTAL	0.07946	Z-Score Abs. Difference Court and Lower Ct Direction [FE]	0.01150
		TOTAL	0.23391

political party of the judge, and the correlation persists after controlling for available characteristics of the lawyer and the case as well as the best prediction of Supreme Court votes. Here, we add the predicted voice trait ratings over a longer time frame in the prediction model of Supreme Court decisions. A binary outcome is constructed based on whether the justice reversed or affirmed the lower court opinion. Katz, Bommarito, and Blackmun’s model uses a large number of judge and case characteristic features, as well as Court trend and lower court trend features. However, their model does not include advocate audio features.

Establishing causality is beyond the scope of the current paper. The ideal experiment would be to randomize the voice of the lawyer unbeknownst to the Justices or the lawyer, and then test for effects out outcome. Although such an experiment is impossible, future clever research designs may find a way to ground strongly supported causal claims. It is worth noting, however, that prior results assuage some of the concern of omitted variables bias: for example, if the vocal cues are correlated with case weakness, then all judges should respond to vocal cues in the same way.

Features and Feature Engineering

Given the availability of the raw audio data, a choice must be made as to whether to use the raw data or the predicted features in the machine learning prediction. For example, in macroeconomic forecasting that relies on principal components or factor analysis the predicted trait is commonly used. In this case the underlying factor driving multiple economic indicators (eigenvectors) is believed to have continuous distribution. Moreover, since the eigenvectors underlying common trait characteristics are likely to be highly correlated, a sparse model like LASSO is less appropriate. Both principal components analysis and regularization approaches aim to reduce dimensionality. However, regularization is a type of supervised learning (built on the assumed relationship between

the outcome and the predictors), whereas principal components analysis is a type of unsupervised learning (considering only the predictors). Given that the goal of our task is to test for the predictive information in audio cues, we opt to use the (predicted) trait features rather than some version of the raw data or unsupervised representations thereof.

The ratings were on a 1-7 scale. Each rater's rating was normalized by subtracting their average rating and dividing by the standard deviation of their ratings (i.e., z-score). The aggregated z-scores corresponding to every lawyer yields a continuous voice trait ratings for every lawyer. Then the z-scores were made binary: if a z-score was positive, it was replaced with 1, if it was negative it was replaced with -1.

Next, every audio clip of a lawyer's opening statements from 1946–2014 was processed into a fixed number of frames and each frame was vectorized into 13 dimensions, following the standard approach used in voice analysis (MFCC) (Ganchev, Fakotakis, and Kokkinakis 2005). A trained random forest classifier model—a weighted non-parametric model that forms weighted predictions based on nearest neighbors—was used to generate the predicted traits. It was most accurate in predicting perceived masculinity (65.79%) while least accurate in predicting perceived trustworthiness (56.02%). The greater predictability in perceived masculinity is consistent with some results reported in Chen, Halberstam, and Yu (2016a), which plays the voice clips backwards and asks raters to rate the backward clips. Among the perceptual questions, ratings for perceived masculinity were most strongly correlated for the forward and backward clips.

These predicted voice trait ratings were appended to the original data set. More specifically, for the audio clips from 1998–2012, the binarized version of the originally obtained continuous z-score ratings were appended and for audio clips from 1946–1997 and 2013–2014, the binary voice traits predicted from the above mentioned model were appended.

Results

The model is evaluated with the binary voice features, which improve case-wise accuracy by 1.1 percentage points from 0.634 to 0.645 and decreases justice-wise accuracy by 0.1 percentage points from 0.649 to 0.648.

The following charts show the feature weights. To present the relative scale, the intercept at the bottom of the figure corresponds to the most important feature present in the model.² Since this is a random forest model, the feature charts do not speak to the directionality of the features' effects.³

An extension of this study can examine the predictive power of other audio features, such as the number of interruptions, the political dialect (Kadiri et al. 2018), vocal implicit bias (Jaiswal et al. 2018), or phonetic accommodation (Chen and Yu 2016). A richer high-dimensional characterization or hand-labeling of the audio clips may also prove fruitful.

Discussion

This chapter examines the possibility that vocal features can influence judicial decisions. In Chen, Halberstam, and Yu (2016a, 2016b) an econometric analysis reveals that vocal features impact court outcomes, even

2. The most important feature is "justice_cumulative_lc_direction_diff" (the difference between the lower court disposition direction and the justice's cumulative direction). "Disposition direction" is a measure of whether the decision of the court whose decision the Supreme Court reviewed was itself liberal or conservative. "Previous" refers to previous Supreme Court term and "cumulative" refers to all prior terms. As such, these two indicators are measurements related to ideology, and in particular, the ideological differences between the Justice and the lower court opinion.

3. Additionally, to address the question of whether the audio features were really picking up lawyer gender, in additional analysis including the gender variables did not increase the accuracy of the model and the average gender scores for petitioner and respondent sides were not in the top 30 most important features.

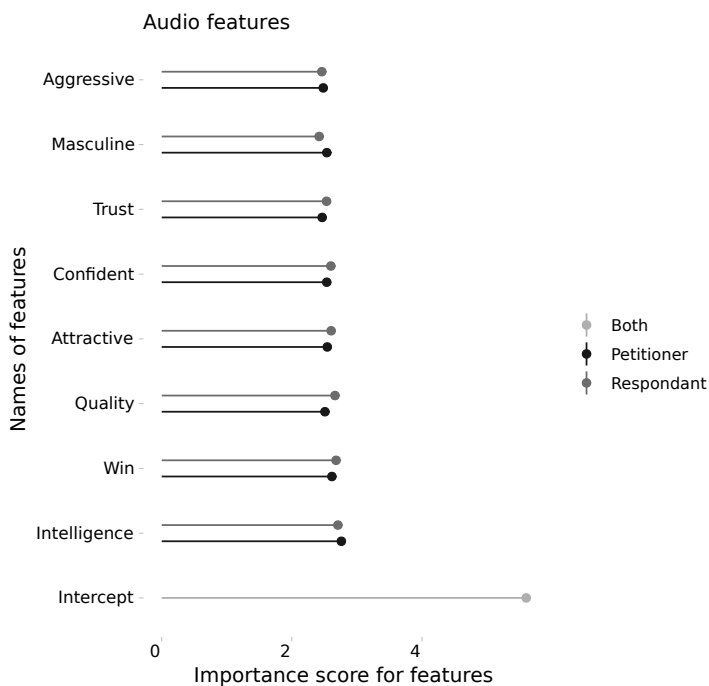


Figure 2. *Feature Weights Relative to Top Feature Weight*

based on a speech sample of less than three seconds. The connection between vocal characteristics and court outcomes was specific only to perceptions of masculinity and not other characteristics. Consistent with employers mistakenly favoring lawyers with masculine voices, perceived masculinity was negatively correlated with winning and the negative correlation was larger in more masculine-sounding industries. The first lawyer to speak was the main driver. Among these petitioners, males below median in masculinity were 7 percentage points more likely to win in the Supreme Court. Republicans, more than Democrats, vote for more feminine-sounding females, while Democrats, but not Republicans, vote

for less masculine-sounding men. Perceived masculinity explains additional variance relative to and is orthogonal to the best random forest prediction model of Supreme Court votes.

In this chapter, we extend this initial work and show that the best predictive model of Supreme Court votes improves with the addition voice characteristics of Supreme Court advocates for almost 70 years of data, rather than 15. The improvement appears robust for predicting Supreme Court case outcomes and appears limited for predicting Supreme Court Justice votes, similar to the finding of Chen, Halberstam, and Yu (2016a). Chen and Kumar (2018) finds that facial features of lawyers also predict Supreme Court votes. A surprising finding across these papers is that advocate characteristics received half as much in importance weight as the most important feature typically attributed to political ideology.

Acknowledgments

We thank Vishal Motwani and Phil Yeres for early assistance. Work on this project was conducted while the first author received financial support from the European Research Council (Grant No. 614708) and Agence Nationale de la Recherche.

References

- Balkin, Jack M. 2011. *Living originalism*. Harvard University Press.
- Ballew, Charles C, and Alexander Todorov. 2007. "Predicting political elections from rapid and unreflective face judgments." *Proceedings of the National Academy of Sciences* 104 (46): 17948–17953.
- Becker, Gary S. 2010. *The economics of discrimination*. University of Chicago press.
- Bertrand, Marianne, and Esther Duflo. 2016. *Field Experiments on Discrimination*. Technical report. National Bureau of Economic Research.
- Cameron, Charles M. 1993. "New Avenues for Modeling Judicial Politics." In *Conference on the Political Economy of Public Law*. Rochester, NY: W. Allen Wallis Institute of Political Economy, University of Rochester. <https://www.princeton.edu/%7B~%7Dccameron/NewAvenuesFigures.pdf>.
- Campbell-Kibler, Kathryn. 2010. "Sociolinguistics and perception." *Language and Linguistics Compass* 4 (6): 377–389.
- Case, Mary Anne C. 1995. "Disaggregating gender from sex and sexual orientation: The effeminate man in the law and feminist jurisprudence." *Yale Law Journal* 105 (1): 1.
- Chen, Daniel L., Yosh Halberstam, and Alan C L Yu. 2016a. "Covering: Mutable Characteristics and Perceptions of (Masculine) Voice in the U.S. Supreme Court."
- . 2016b. "Perceived masculinity predicts US Supreme Court outcomes." *PloS one* 11 (10): e0164324.
- Chen, Daniel L., and Manoj Kumar. 2018. "Is Justice Really Blind? And Is It Also Deaf?"

- Chen, Daniel L., and Alan Yu. 2016. "Mimicry: Phonetic Accommodation Predicts U.S. Supreme Court Votes."
- Eckert, Penelope. 2008. "Variation and the indexical field." *Journal of Sociolinguistics* 12 (4): 453–476.
- Ganchev, Todor, Nikos Fakotakis, and George Kokkinakis. 2005. "Comparative evaluation of various MFCC implementations on the speaker verification task." In *Proc. of the SPECOM-2005*, 191–194.
- Goffman, Erving. 1963. *Stigma: Notes on the management of spoiled identity*. NJ: Prentice-Hall.
- Grogger, Jeffrey. 2011. "Speech Patterns and Racial Wage Inequality." *The Journal of Human Resources* 46 (1): 1–25.
- Jaiswal, Divisha, Hitesh Ochani, Ramya Vunikili, Richa Deshmukh, Daniel L. Chen, and Elliott Ash. 2018. "Vocal Implicit Bias." *Proceedings of Experimental Linguistics*.
- Kadiri, Yassine, Thomas Leble, Zsolt Pajor-Gyulai, Elliott Ash, and Daniel L. Chen. 2018. "Tone of Voice Predicts Political Attitudes: Evidence from U.S. Supreme Court Oral Arguments."
- Katz, Daniel Martin, Michael James Bommarito, and Josh Blackman. 2014. "Predicting the behavior of the {S}upreme {C}ourt of the {U}nited {S}tates: A general approach." *Available at SSRN* 2463244.
- Klofstad, Casey A, Rindy C Anderson, and Susan Peters. 2012. "Sounds like a winner: voice pitch influences perception of leadership capacity in both men and women." *Proceedings of the Royal Society of London B: Biological Sciences* 279 (1738): 2698–2704.
- Korn, Jane. 2004. "Teaching talking: Oral communication skills in a law course." *Journal of Legal Education* 54:588.

- Kornhauser, Lewis A. 1999. "Judicial Organization and Administration." Chap. 11 in *Encyclopedia of Law and Economics*, edited by Chris William Sanchirico, 5:27–44.
- Mayew, William J, and Mohan Venkatachalam. 2012. "The Power of Voice: Managerial Affective States and Future Firm Performance." *Journal of Finance* 67 (1): 1–43.
- McAleer, Phil, Alexander Todorov, and Pascal Belin. 2014. "How do you say hello? {P}ersonality impressions from brief novel voices." *PLoS ONE* 9 (3): e90779.
- Nass, Clifford, and Kwan Min Lee. 2001. "Does computer-synthesized speech manifest personality? experimental tests of recognition, similarity-attraction, and consistency-attraction." *Journal of Experimental Psychology: Applied* 7:171–181.
- Neumark, David. 2016. *Experimental Research on Labor Market Discrimination*. Technical report. National Bureau of Economic Research.
- Posner, Richard A. 1973. "An Economic Approach to Legal Procedure and Judicial Administration." *The Journal of Legal Studies* 2 (2): 399–458.
- Purnell, Thomas, William Idsardi, and John Baugh. 1999. "Perceptual and Phonetic Experiments on {A}merican {E}nglish Dialect Identification." *Journal of Language and Social Psychology* 18 (10): 10–30.
- Scherer, Klaus R. 1979. "Voice and speech correlates of perceived social influence in simulated juries." In *The social psychology of language*, edited by Klaus R Scherer, Howard Giles, and R St. Clair, 88–120. London: Blackwell.
- Starecheski, Laura. 2014. *Can Changing How You Sound Help You Find Your Voice?* <http://www.npr.org/blogs/health/2014/10/14/354858420/can-changing-how-you-sound-help-you-find-your-voice>.

- Tigue, Cara C, Diana J Borak, Jillian J M O'Connor, Charles Schandl, and David R Feinberg. 2012. "Voice pitch influences voting behavior." *Evolution and Human Behavior* 33:210–216.
- Todorov, Alexander, Anesu N Mandisodza, Amir Goren, and Crystal C Hall. 2005. "Inferences of competence from faces predict election outcomes." *Science* 308 (5728): 1623–1626.
- Yoshino, Kenji. 2006. *The Hidden Assault on Our Civil Rights*. Random House Incorporated.