A Data Collection Process for Worker Samples

A.1 Main Dataset: Lawyers from Vault 100 Law Firms

The Vault 100 is an annual ranking of the most prestigious law firms in the U.S. I consider law firms that were listed in the Vault 100 rankings at least once between 2016 and 2018. Based on information from a pilot sample, I dropped firms that had a live receptionist 24/7 and firms where less than 10 percent of voicemail greetings were self-recorded by lawyers. This left me with 84 law firms. Table S1 provides a list of these law firms.

From this list of 84 top private law firms, I used web scraping to collect the directory of lawyers' names and phone numbers. Starting with pilot calls in May 2017, and mostly in late December 2017 and early January 2018, I collected 57,064 distinct phone numbers and used Voicent, an automated phone-calling software, to call each number during non-working hours (typically 2-5 AM EST), and to record each call that was successfully connected. I recorded the first 10 seconds of each successful call and extracted the first 3 seconds of each recording to minimize the likelihood of capturing silence or machine-generated audio, such as generic instructions for leaving a message.

Table S1. List of law firms.

Akin Gump Strauss Hauer & Feld LLP	Keker, Van Nest & Peters LLP
Allen & Overy LLP	Kilpatrick Townsend & Stockton LLP
Alston & Bird LLP	King & Spalding LLP
Arent Fox LLP	Kirkland & Ellis
Arnold & Porter Kaye Scholer LLP	Kramer Levin Naftalis & Frankel LLP
BakerHostetler	Latham & Watkins LLP
Baker McKenzie	Locke Lord LLP
Ballard Spahr LLP	Manatt, Phelps & Phillips, LLP
Blank Rome LLP	Mayer Brown LLP
Boies Schiller Flexner LLP	McDermott Will & Emery
Bracewell LLP	Milbank LLP
Bryan Cave Leighton Paisner LLP	Morgan, Lewis & Bockius LLP
Cadwalader, Wickersham & Taft LLP	Morrison & Foerster LLP
Cahill Gordon & Reindel LLP	Nixon Peabody LLP
Cleary Gottlieb Steen & Hamilton LLP	Norton Rose Fulbright LLP
Clifford Chance US LLP	O'Melveny & Myers LLP
Covington & Burling LLP	Patterson Belknap Webb & Tyler LLP
Crowell & Moring LLP	Paul Hastings LLP
DLA Piper	Paul, Weiss, Rifkind, Wharton & Garrison LLP
Davis Polk & Wardwell LLP	Pepper Hamilton LLP
Davis Wright Tremaine LLP	Perkins Coie LLP
Debevoise & Plimpton LLP	Pillsbury Winthrop Shaw Pittman LLP
Dechert LLP	Proskauer Rose LLP
Dentons	Ropes & Gray LLP
Dorsey & Whitney LLP	Schulte Roth & Zabel LLP
Faegre Drinker Biddle & Reath LLP	Seyfarth Shaw LLP
Fenwick & West LLP	Shearman & Sterling
Finnegan, Henderson, Farabow, Garrett & Dunner, LLP	Sheppard Mullin
Fish & Richardson P.C.	Sidley Austin LLP
Foley & Lardner LLP	Simpson Thacher & Bartlett LLP
Fox Rothschild LLP	Skadden, Arps, Slate, Meagher & Flom LLP and Affiliates
Gibson, Dunn & Crutcher LLP	Squire Patton Boggs
Goodwin Procter LLP	Steptoe & Johnson LLP
Haynes and Boone, LLP	Susman Godfrey LLP
Holland & Knight LLP	Troutman Sanders LLP
Hunton & Williams LLP	Venable LLP
Irell & Manella LLP	Vinson & Elkins LLP
Jenner & Block LLP	Weil, Gotshal & Manges LLP
Jones Day	White & Case LLP
K&L Gates LLP	Williams & Connolly LLP
Kasowitz Benson Torres LLP	Wilson Sonsini Goodrich & Rosati
Katten Muchin Rosenman LLP	Winston & Strawn LLP

To detect poor quality clips, I decomposed each 3-second clip into 225 subintervals. Specifically, each clip is represented as a time interval [0, 3] in seconds, and each subinterval in that clip is defined as $\left[\frac{3k}{225}, \frac{3(k+1)}{225}\right]$ for $k \in \{0, \ldots, 224\}$. I used Praat's "To Intensity" function on the upper bound of each subinterval. This function, which estimates the amplitude at a given point, returns the value -300 dB if the clip is silent. For each clip, if more than 30% of the 225 sample points had an intensity of -300 dB, I eliminated the clip from my sample. A substantial number of clips were of poor quality, for example, due to unnatural acceleration or fragmentation of sounds in the clip. These issues often resulted in a high proportion of silence in the clip.

After eliminating unsuccessful recordings and poor quality clips, I extracted 39,962 lawyers' voicemail recordings from the 57,064 phone numbers. These comprise the main dataset of lawyers.

A.2 Verified Female Lawyers from the Main Dataset

To determine which clips from the main dataset were self-recorded by a female lawyer, I listened to all clips from lawyers who were classified as female by ALM. If a clip was entirely recorded in first person by a human speaker, I classified the clip as self-recorded. This is in contrast to (1) assistant-recorded clips, which are in third-person, (2) ambiguous clips, which are human-recorded but do not contain first-person or third-person pronouns, (3) machine-recorded clips, which are clearly distinct from human voices, and (4) combinations of human-recorded and machine-recorded clips. I further eliminated poor-quality clips, for instance, clips that contained static, live answers, dialing noises, or unnatural acceleration of sounds. Through this process, I identified 6,618 self-recorded clips; however, 210 of them sounded male, and further information from their webpages, such as profile pictures and gender pronouns, confirmed this for 209 of them. I was unable to find gender information on one lawyer and omitted the clip from analysis.

Given these verification procedures, I decided to use the sample of 6,408 clips. The finite mixture models failed to converge on a solution for 9 of these clips, leaving me with a final sample of 6,399 clips that were verified as self-recorded by a female lawyer.

A.3 Auxiliary Dataset: Law Firm Assistants

To find recordings of executive assistants of female lawyers, I listened to all clips from lawyers who were classified as female by ALM. To maximize the chances of finding recordings of female assistants of male lawyers, I listened to all clips which fulfilled all the following criteria:

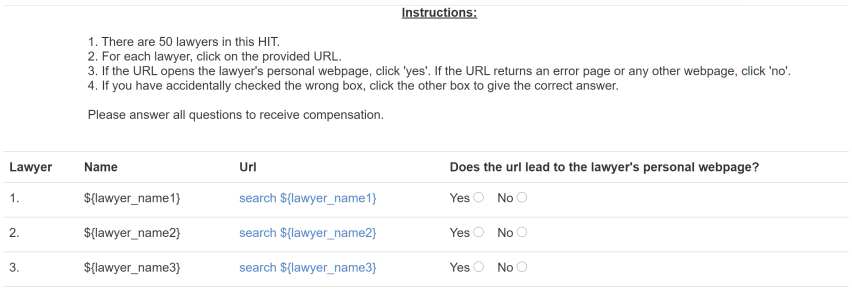
1. The lawyers associated with these clips were classified as male by ALM.
2. The clips had a mean frequency above 150 Hz, well above the unitary male frequency mode of 100 Hz.
3. The clips had a probability of being self-recorded of at least 0.25, as determined by my machine learning classification (see Section C for details on the machine learning classification process).

I classified a clip as assistant-recorded if the clip contained third-person pronouns and was recorded by a human. I eliminated clips that were recorded by non-executive assistants (i.e., the assistants who spoke generally on behalf of the firm rather than for a specific lawyer), clips that were fully or partially machine-recorded, and poor-quality clips. Through this process, I found 237 clips that were fully recorded by female executive assistants on behalf of female lawyers, and 412 clips that were fully recorded by female executive assistants on behalf of male lawyers.

A.4 Auxiliary Dataset: Lawyers Who Switched Firms

I define "switchers" as lawyers who had switched firms since the initial collection of their voicemail greeting. As noted in Section A.1, the period of initial collection was from May 2017 (pilot calls) to January 2018, with most calls in late December 2017 and early January 2018. To find the switchers, about two years after the period of initial collection, I recruited MTurk workers to check the webpages of lawyers whose voicemails had a probability of being self-recorded of at least 0.5, as determined by my machine learning classification (see Section C for details on the machine learning classification process). The MTurk workers were selected on a first-come-first-served basis from a pool of US-based MTurk workers who had a HIT approval rate of 99% or more and who had over 10,000 approved HITs.

I determined whether a lawyer had switched firms by clicking on their original profile page URL. If the URL linked to the lawyer's profile page, I determined that the lawyer had not switched firms. If the URL did not link to the lawyer's profile page, for instance by redirecting to a general directory or an error page, I determined that the lawyer had switched firms. See Figure S1 for the survey format.



Instructions:

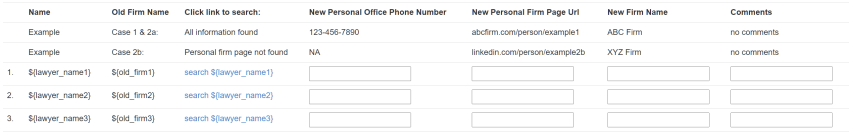
1. There are 50 lawyers in this HIT.
2. For each lawyer, click on the provided URL.
3. If the URL opens the lawyer's personal webpage, click 'yes'. If the URL returns an error page or any other webpage, click 'no'.
4. If you have accidentally checked the wrong box, click the other box to give the correct answer.

Please answer all questions to receive compensation.

Lawyer	Name	Url	Does the url lead to the lawyer's personal webpage?
1.	${lawyer_name1}	search ${lawyer_name1}	Yes ○ No ○
2.	${lawyer_name2}	search ${lawyer_name2}	Yes ○ No ○
3.	${lawyer_name3}	search ${lawyer_name3}	Yes ○ No ○

Figure S1. MTurk HIT for finding lawyers who switched firms.

I then recruited MTurk workers to find the switchers' new firms, profile page URLs, and personal office phone numbers. I provided MTurk workers with the switchers' names and old firms, and instructed respondents not to use sites other than Google, LinkedIn, and the lawyer's new firm website when providing information to avoid third-party websites of unknown reliability. See Figure S2 for details on the survey format and instructions.



	Name	Old Firm Name	Click link to search:	New Personal Office Phone Number	New Personal Firm Page Url	New Firm Name	Comments
	Example	Case 1 & 2a:	All information found	123-456-7890	abcfirm.com/person/example1	ABC Firm	no comments
	Example	Case 2b:	Personal firm page not found	NA	linkedin.com/person/example2b	XYZ Firm	no comments
1.	${lawyer_name1}	${old_firm1}	search ${lawyer_name1}				
2.	${lawyer_name2}	${old_firm2}	search ${lawyer_name2}				
3.	${lawyer_name3}	${old_firm3}	search ${lawyer_name3}				

Figure S2. MTurk HIT for finding new office numbers.

As in Section A.1, I used Voicent to collect voicemails from the new phone numbers. For each lawyer, I checked if both the old and new recordings were self-recorded by the same lawyer following the procedure in Section A.2. In total, I found 627 lawyers with self-recorded voicemails at both their old and new firms. Of these 627 lawyers, 198 lawyers were female, as further verified using the procedure described above.

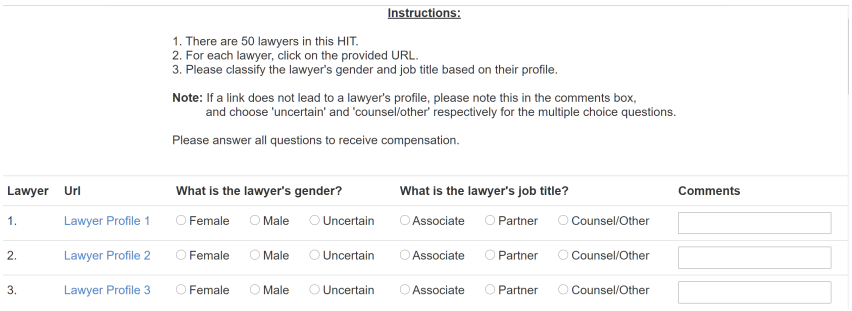
A.5 Auxiliary Dataset: Promoted Lawyers

About two years after the period of initial data collection, I recruited MTurk respondents to check the job title of a subsample of lawyers from the lawyers' webpages. The workers were selected on a first-come-first-served basis from a pool of US-based MTurk workers who had a HIT approval rate

of 99% or more and who had over 10,000 approved HITs. I selected lawyers who fulfilled all the following criteria:

1. Confirmed to have stayed at their original firm.
2. Associates as of May 2017 – January 2018.
3. Had voicemails with a probability of being self-recorded of at least 0.5 (see Section C for details on the machine learning classification process).

I provided MTurk workers with the lawyers' website URLs, and instructed them to classify the lawyers' job title based on their personal webpage profiles. See Figure S3 for the survey format. This process elicited promotions data for 1,925 female lawyers whose clips were verified as self-recorded by a female lawyer using the procedure described earlier, and who were Associates as of May 2017 – January 2018. Of these 1,925 lawyers, 196 were promoted to Partner and 137 were promoted to Counsel/Other within the same firm as of January 2020.



Instructions:

1. There are 50 lawyers in this HIT.
2. For each lawyer, click on the provided URL.
3. Please classify the lawyer's gender and job title based on their profile.

Note: If a link does not lead to a lawyer's profile, please note this in the comments box, and choose 'uncertain' and 'counsel/other' respectively for the multiple choice questions.

Please answer all questions to receive compensation.

Lawyer	Url	What is the lawyer's gender?			What is the lawyer's job title?			Comments
1.	Lawyer Profile 1	○ Female	○ Male	○ Uncertain	○ Associate	○ Partner	○ Counsel/Other	
2.	Lawyer Profile 2	○ Female	○ Male	○ Uncertain	○ Associate	○ Partner	○ Counsel/Other	
3.	Lawyer Profile 3	○ Female	○ Male	○ Uncertain	○ Associate	○ Partner	○ Counsel/Other	

Figure S3. MTurk HIT for finding promoted lawyers.

A.6 Auxiliary Dataset: Supreme Court Lawyers

From www.oyez.org, I collected data from 129 oral arguments made by female advocates at the U.S. Supreme Court between 1985 and 2005. From the recording of each argument, I took three voice samples and extracted the first 3 seconds of each sample. The samples are from the opening sentence, closing sentence, and one sentence taken from the middle of the argument (approximately minute 15).

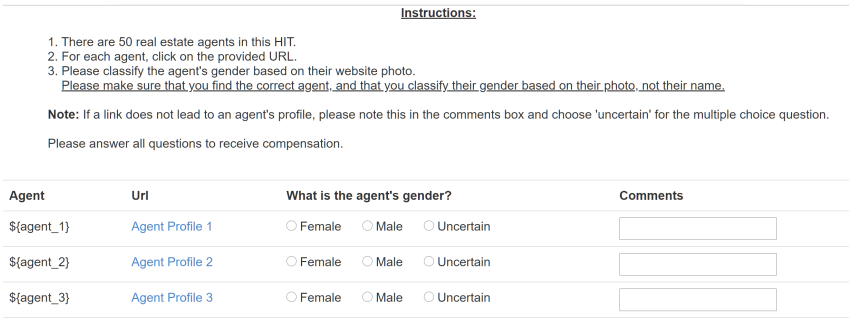
A.7 Auxiliary Dataset: RE/MAX Real Estate Agents

I collected the name, franchise, directory URL, and office number of 1,694 RE/MAX residential real estate agents through web scraping. These agents constituted all agents in Chicago, Dallas, Houston, and Phoenix in the fields of first-time buyers, luxury properties, condominiums, residential acreages, and rentals that were listed on www.remax.com. I also collected the name, franchise, directory URL, and office number of 2,013 RE/MAX commercial real estate agents listed on http://www.remaxcommercial.com/Roster/Agents through web scraping.

I used Voicent to collect voicemails from these phone numbers, and manually classified these clips as self-recorded by the agent by listening to them. I successfully collected 539 self-recorded voicemails from the residential real estate agents, and 527 self-recorded voicemails from the commercial real estate agents.

I then recruited MTurk respondents to check the gender of agents who had self-recorded voicemails. They were selected on a first-come-first-served basis from a pool of US-based MTurk workers who had a HIT approval rate of 99% or more and who had over 10,000 approved HITs. I provided MTurk

respondents with the agents' names and directory URLs, and instructed them to classify the agents' gender based on their profile pictures. See Figure S4 for the survey format. Of the 539 residential real estate agents, 337 were classified as female. Of the 527 commercial real estate agents, 159 were classified as female.



Instructions:

1. There are 50 real estate agents in this HIT.
2. For each agent, click on the provided URL.
3. Please classify the agent's gender based on their website photo.
   Please make sure that you find the correct agent, and that you classify their gender based on their photo, not their name.

Note: If a link does not lead to an agent's profile, please note this in the comments box and choose 'uncertain' for the multiple choice question.

Please answer all questions to receive compensation.

Agent	Url	What is the agent's gender?	Comments
${agent_1}	Agent Profile 1	○ Female ○ Male ○ Uncertain	
${agent_2}	Agent Profile 2	○ Female ○ Male ○ Uncertain	
${agent_3}	Agent Profile 3	○ Female ○ Male ○ Uncertain	

Figure S4. MTurk HIT for real estate agents.

B Estimating Frequencies from Audio Data

I used Praat, an open-source program to extract frequency data from each clip (Boersma, 1993). Praat chooses the frequency candidate associated with the highest *local* strength subject to various thresholds and the global path finder, a system that penalizes frequency variation across adjacent frames. This way, background noise and nonhuman sounds are less likely to confound estimates.

Specifically, to extract frequency data from each clip, I used the following baseline parameters for the function "To Pitch (ac)":

Pitch Floor: 50 Hz;
Pitch Ceiling: 400 Hz;
Time Step: 5 milliseconds;
Window: Hanning;
Silence Threshold: 0.03;
Voicing Threshold: 0.45;
Octave Cost: 0.01;
Octave-Jump Cost: 0.35;
Voiced/Unvoiced Cost: 0.14.

These parameters are defined by Praat as follows (Boersma, 1993; Boersma, 2001):

- *Pitch Floor:* The minimum frequency that will be considered for estimation. A pitch floor of 50 Hz is well below the range of human voice frequencies produced by the natural voice register. The pitch floor also determines the length of the analysis window, which is 3/*Pitch Floor* seconds long. With a pitch floor of 50 Hz, the analysis window is 0.06 seconds long.
- *Pitch Ceiling*: The maximum frequency that will be considered for estimation. A pitch ceiling of 400 Hz is well above the range of human voice frequencies produced by the natural voice register.
- *Time Step*: The interval between frequency estimates. The points in the clip where the first and the last frequency estimates are taken depend on the length of the analysis window, which in turn depends on the pitch floor. For a clip represented as a time interval [0, 3] in seconds, a pitch floor of 50 Hz and time step of 5 milliseconds mean that Praat produces one frequency

estimate per 5 milliseconds in the subinterval [0.03, 2.97], for a total of 589 frequency estimates including the endpoints of the subinterval.

I used the default Praat settings for the following:

- *Window*: I use the default Hanning window.
- *Silence Threshold*: For each frame, if the local absolute amplitude peak is less than approximately Silence Threshold times the global absolute amplitude peak, the frame will be classified as "voiceless" (the frequency estimate will be a missing value). I use the default 0.03 as the silence threshold.
- *Voicing Threshold*: For each frame, if the strengths of all frequency candidates in the frame are below Voicing Threshold, the frame will be classified as "voiceless" (the frequency estimate will be a missing value). I use the default 0.45 as the voicing threshold.
- *Octave Cost*: This parameter determines how much higher-frequency candidates are favored relative to lower-frequency candidates. It is necessary to force Praat to choose a frequency candidate in the case of a perfectly periodic signal, where all autocorrelation peaks have equal values. I use the default 0.01 per octave as the octave cost.
- *Octave-Jump Cost*: This parameter determines the extent to which rapid pitch changes between adjacent frames are disfavored. In conjunction with Voiced/Unvoiced Cost, this is a global path finder parameter that affects estimates across rather than only within frames. I use the default 0.35 as the octave-jump cost.
- *Voiced/Unvoiced Cost*: This parameter determines the extent to which rapid transitions between voiced and voiceless frames are disfavored. In conjunction with Octave-Jump Cost, this is a global path finder parameter that affects estimates across rather than only within frames. I use the default 0.14 as the voiced/unvoiced cost.

With these parameters, each clip comprises 589 frames. For each frame, the best frequency candidates were determined using the function "Get value in frame". Through this process, I obtained 589 frequency estimates for each clip, though many of these estimates are voiceless (i.e. missing values), and the number of such missing values differs across clips. See Table S2 for robustness checks on the values of the parameters used. Figure S5 shows these estimates over time for two greetings with the same average pitch. Figure S6 repeats the estimation on a pilot sample with the pitch floor lowered to 25 Hz, to check whether the low band reflects vocal fry.

C Machine Learning Classification Process

I define a self-recorded clip as a clip which contains only a lawyer's voice, as opposed to an automated voice, an assistant's voice, or a combination of voices. Due to the large number of clips in the main dataset of 39,962 lawyers, I used machine learning to predict the probability that a clip is self-recorded for the main dataset. I used text information, acoustic information, and demographic information as predictors.

I obtained text information for each clip by transcribing the clips with IBM Watson Speech Recognition API. I decomposed the transcribed text into individual words, and selected the 50 most frequent words and the number of words per clip as predictors. I further manually selected 12 frequently occurring phrase patterns as predictors.

I obtained acoustic information for each clip using Praat and selected eight acoustic variables as predictors. I also obtained the demographic information of each lawyer from their websites. This information included: job title, practice area, law school, undergraduate school, any graduate degrees, graduation year for each degree earned, academic honors earned, and gender. Gender was assessed

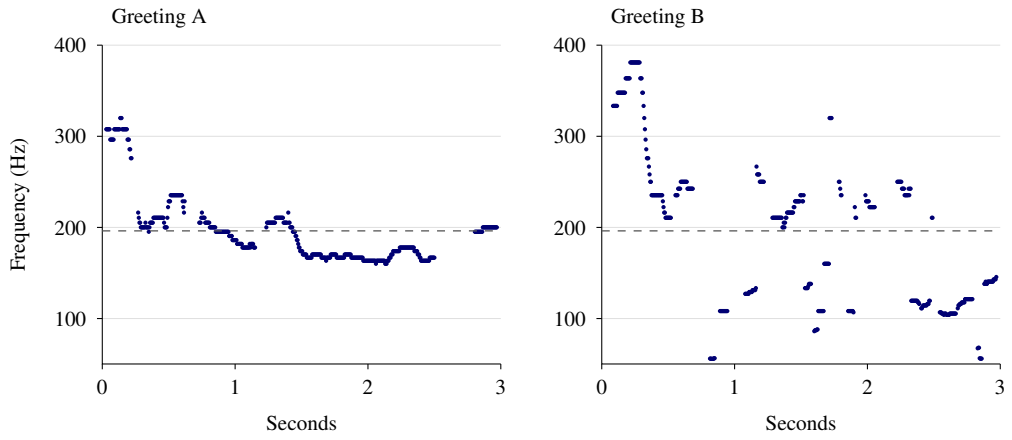


Figure S5. Same average pitch, different patterns.
Notes: Frequency estimates every 5 milliseconds for two voicemail greetings with similar averages, drawn at random from the random sample of 1,000 clips of verified self-recorded greetings by female lawyers (Table 5), and shown as an illustration. Estimates use the baseline parameters in Section B. The dashed line marks each greeting's average; both average 196 Hz. Greeting A stays in one band around its average. Greeting B moves between a band near 200 Hz and a band near 100 Hz.

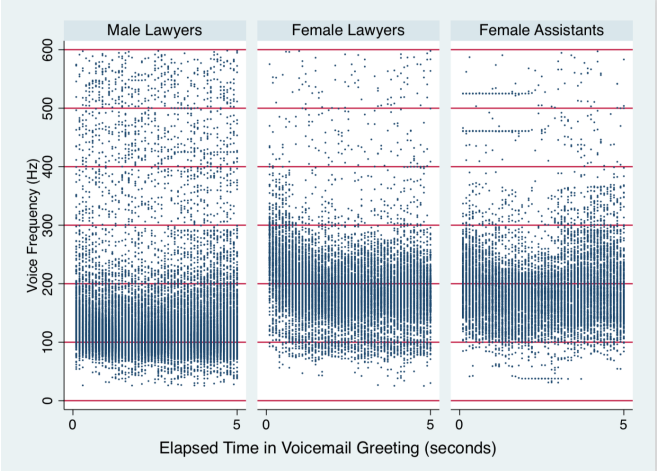


Figure S6. Vocal fry check: voice frequency over time with a 25 Hz floor.
Notes: Voice frequencies from a pilot sample of voicemail greetings recorded by male lawyers (n = 813), female lawyers (n = 364) and female assistants (n = 388). The data are 50 frequency samples per greeting, taken at 0.1-second intervals over the first five seconds; each dot is a speaker's mean frequency in an interval. The pitch floor is lowered from the baseline 50 Hz to 25 Hz (ceiling 600 Hz), so that vocal fry, which averages about 50 Hz and typically comes at the end of a sentence (Blomgren et al., 1998; Wolk et al., 2012), would appear if present. For female lawyers, the low band sits near 100 Hz, not 50 Hz, and appears throughout the greeting, not only at its end.

Table S2. Frequency estimation robustness checks.

Row	Frequency Estimates	Group	95% C.I. (Share)	95% C.I. (Location)	Group	95% C.I. (Share)	95% C.I. (Location)	Obs.
1	Residualized Baseline	Two-mode	(0.35, 0.38)	(−48.24, −46.33)	One-mode	(0.62, 0.65)	(25.94, 28.04)	4,682
2	No Pathfinder	Two-mode	(0.61, 0.67)	(80.94, 83.02)	One-mode	(0.33, 0.39)	(158.86, 164.57)	999
3	Stronger Pathfinder	Two-mode	(0.31, 0.38)	(83.36, 87.65)	One-mode	(0.62, 0.69)	(160.38, 164.32)	998
4	Double OctaveCost	Two-mode	(0.31, 0.37)	(83.71, 87.76)	One-mode	(0.63, 0.69)	(160.62, 164.60)	997
5	Gaussian Window	Two-mode	(0.33, 0.40)	(84.47, 88.31)	One-mode	(0.60, 0.67)	(160.95, 164.89)	998
6	60 Hz Floor	Two-mode	(0.34, 0.41)	(84.67, 87.71)	One-mode	(0.59, 0.66)	(157.86, 162.42)	997
7	40 Hz Floor	Two-mode	(0.30, 0.36)	(84.19, 88.39)	One-mode	(0.64, 0.70)	(161.03, 164.88)	988

Notes: The first row shows results using the residuals from regressing the location of the low mode (estimates from baseline 5-comp. FMM) on years of experience, firm, title, and litigator fixed effects. All subsequent rows use the baseline FMM. Rows 2-7 use the same random subset of 1,000 recordings used in Tables 5 and 6. The second row switches off the pathfinder feature in Praat, which penalizes sharp changes in frequency. The location estimates are similar, but the distribution of groups is significantly different. The third row uses pathfinder parameters with increased values, which is the opposite of switching off the pathfinder. The fourth row uses double the *OctaveCost* as the baseline specification, thereby favoring higher-frequency candidates. The fifth row uses a Gaussian window to smooth the data rather than the default Hanning window. The sixth row uses a higher pitch floor, which implies a narrower analysis window of 50 milliseconds, and the seventh row uses a lower pitch floor, which implies a wider analysis window of 75 milliseconds.

by scrutinizing lawyers' first names, subjectively classifying photos, searching for gendered pronouns in the lawyers' biographical descriptions, and cross-referencing the recorded greeting. Generally, these gender indicators perfectly corroborated one another.

I set aside about 10% of the sample (the "ML sample") for training, testing, and validation. I manually classified all clips in the ML sample as self-recorded or otherwise following the procedure above. I randomly selected 80% of the ML sample for training and validation (the "training and validation sample") and the remaining 20% for testing ("the testing sample"). To address any possible overfitting issues, I used 5-fold cross validation to tune the model with the objective of minimizing logarithmic loss. I applied this procedure to four machine learning models: random forest, support vector machine, k-nearest neighbors, and XGBoost. XGBoost significantly outperformed the other three models with a predictive accuracy of 93.52%. With a probability threshold of 0.95, the XGBoost model had a predictive accuracy of 99.12%. I thus used XGBoost as my final machine learning model.

The most important non-acoustic variable was the number of words spoken by the lawyer in the greeting (coming in at sixth place). Gender was the most important, and indeed the only, demographic variable identified in the top-fifteen most predictive characteristics (coming in at eighth place).

D Survey Design for Relative Perceptions

To check if bimodality is correlated with perceptions of lawyers' characteristics, I surveyed 200 MTurk workers and asked them to listen to pairs drawn from the same set of 200 matched pairs (each rater heard a random 10). After listening to each pair, workers were asked to rate the speakers on a relative scale on competitiveness, dominance, risk-taking attitude, seniority, and trustworthiness on a seven-point Likert scale. The survey instructions and format are shown in Figure S7.

For each question, the position of the bimodal clip was randomized as either Recording 1 or Recording 2, and across questions for each HIT, the bimodal clip appeared 5 times as Recording 1 and 5 times as Recording 2. The order of characteristics on the Likert scale was also randomized for each question. These specifications, as well as any information about bimodality, were not made known to the workers.

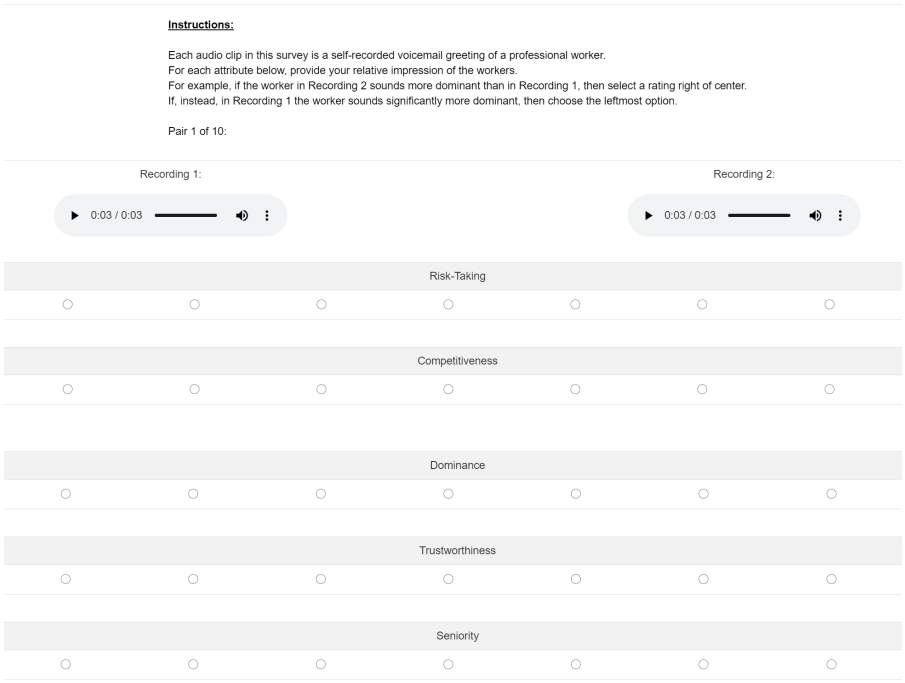


Figure S7. MTurk HIT for relative attribute ratings.

Respondents were selected on a first-come-first-served basis from a pool of US-based MTurk workers who had a HIT approval rate of 99% or more and who had over 10,000 approved HITs. Only one HIT per respondent was considered; if the respondent completed more than one HIT, all additional HITs were rejected. I created male-only and female-only versions of the survey to achieve balance in respondent sex. In all, I had 100 male respondents and 100 female respondents across 200 HITs, with respondents self-reporting their sex.

E Human Detection of Two-Mode Voices

Can listeners consciously tell two-mode and one-mode voices apart? The brain responds to brief and subtle differences in sound, but that does not mean people can hear them consciously. To test this, I recruited 100 female and 100 male workers on Amazon's Mechanical Turk (MTurk). Recruitment was first-come, first-served and limited to U.S. residents who had completed at least 10,000 tasks on the platform with an approval rating of 99 percent or more. Workers ranged in age from 18 to 74, with a median of 36.

I paired 250 two-mode clips with 250 one-mode clips, so that the two clips in each pair had nearly identical mean pitch (within 1 Hz). Each worker was assigned 50 pairs and asked to identify the two-mode clip in each. The selection of pairs, their order, and the order of the clips within each pair were randomized across workers. To remove any influence of what was said, the clips were played in reverse, which keeps pitch and voice quality while removing the words. Workers practiced on three pairs first, received immediate feedback after each answer along with their running success rate, and earned a bonus for classifying more than half of the pairs correctly. Guessing, or giving the same answer throughout, yields a success rate of 0.5.

Listeners did only slightly better than chance, with a mean success rate of 0.525 (top panel of Figure S8). Men and women performed alike, as did workers above and below the median age. Workers who spent more than the median of 33 minutes on the task did about 1 percentage point better, a

difference that is not statistically significant. Performance varied widely across workers: average success rates by quartile were 0.62, 0.55, 0.50 and 0.43, nearly identical for men and women.

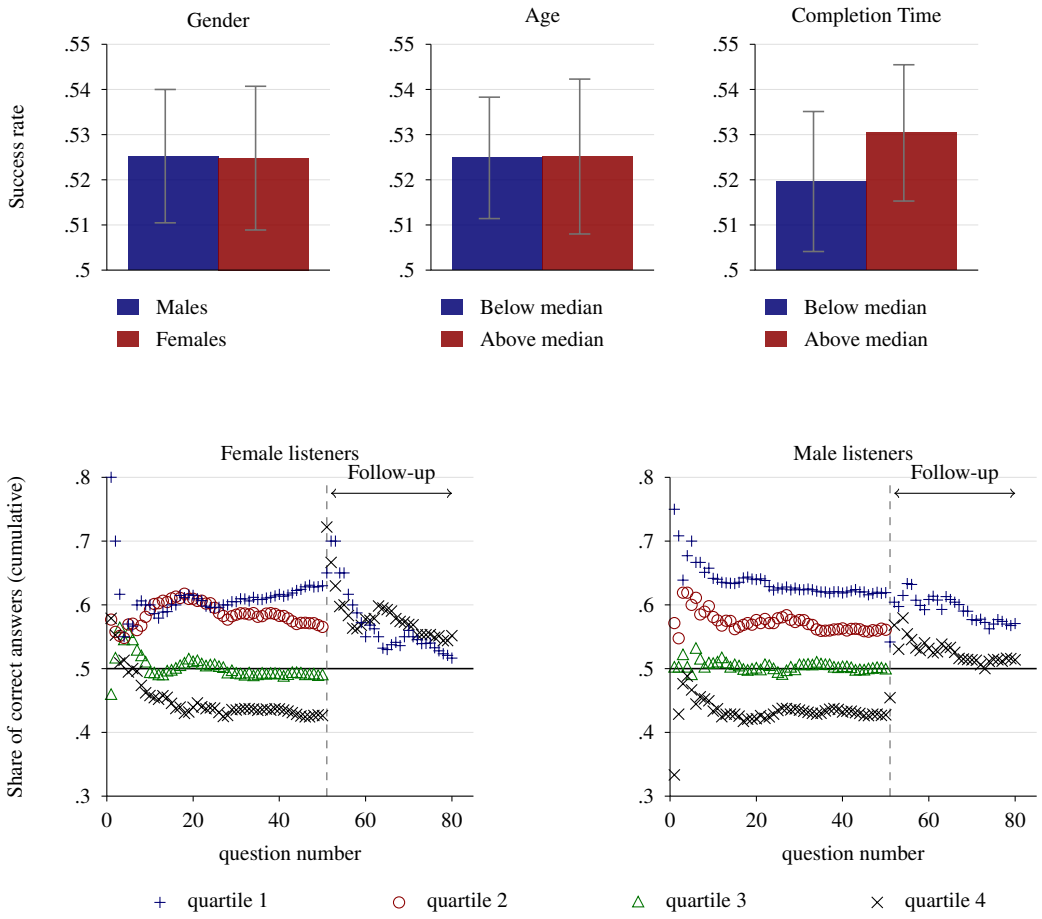


Figure S8. Human detection of two-mode voices.
Notes: 100 female and 100 male U.S. participants recruited on MTurk each classified 50 pairs of clips, one two-mode and one one-mode, with mean pitch within 1 Hz, played in reverse to remove verbal content. Top panel: mean success rate and 95 percent confidence interval by listener gender, age (below or above the median) and completion time (below or above the median). A success rate above 0.5 means listeners did better than chance. Bottom panel: cumulative share of correct answers by question for each quartile of listeners, ranked by performance in the first survey. Questions 51 to 80 are the follow-up survey for the top (quartile 1) and bottom (quartile 4) quartiles.

To see whether listeners can learn to detect the difference, and whether early success holds up, I invited workers in the top and bottom quartiles to a shorter follow-up survey of 30 pairs in the same format. Of these 100 workers, 84 took part (39 women and 45 men). The bottom quartile improved and finished above 0.5, the top quartile partly reverted toward the mean, and all groups did better than chance (bottom panel of Figure S8). Among women, the quartiles separated only about halfway through the first survey; among men, they separated within the first few questions.

In sum, the difference between two-mode and one-mode voices is hard to hear consciously, though some listeners can learn to detect it. Yet in the relative ratings described in Section D, listeners rate the two groups differently.

Table S3. Verbal content comparison.

	Group	Mean	Standard Deviation	Group	Mean	Standard Deviation
Total words	Two-mode	10.563	1.902	One-mode	10.410	2.075
First word *hi* or *hello*	Two-mode	0.505	0.501	One-mode	0.510	0.500
Includes *reached*	Two-mode	0.063	0.244	One-mode	0.062	0.241
Includes *at*	Two-mode	0.042	0.201	One-mode	0.038	0.192
Includes *you've*	Two-mode	0.042	0.201	One-mode	0.044	0.204
Includes *hi*	Two-mode	0.035	0.184	One-mode	0.036	0.186
Includes *I'm*	Two-mode	0.029	0.168	One-mode	0.032	0.175
Unique words within group	Two-mode	0.248	0.432	One-mode	0.216	0.411

Notes: The table uses a random sample of 1,000 clips from the main sample of verified self-recorded greetings by female lawyers (6,399 clips). Among these 3-second clips, 380 are classified as two-mode and 620 as one-mode. The verbal content from each clip was extracted using an automated transcription program. Each row, except for the last, represents summary statistics for each group of clips. *Hi* or *hello* are the most frequently used first word in the greetings of each group. The subsequent five words, *reached*, *at*, *you've*, *hi*, and *I'm*, are the most commonly used words in the greetings of each group. Unique words refers to the share of words that appear only once within a group.

Table S4. Number of recordings in each subsample of Figures 5 and 6.

Figure 5: main sample			Figure 6: auxiliary samples		
Group	Subsample	Recordings	Group	Subsample	Recordings
Practice Area	Litigation	2,924	Promotions	Partner	196
	Other	3,475		Counsel/Other	137
Lawyer Title	Partner	2,281	Separations	Former Firm	198
	Counsel/Other	1,231		Current Firm	198
	Associate	2,887	Oral Arguments	Beginning	129
Vault Law 100	Top Ranked	3,580		Middle	129
	Bottom Ranked	2,819		End	129
Women in Law	Top Ranked	3,351	Law Firm Assistants	Female Employer	237
	Bottom Ranked	2,911		Male Employer	412
Law Firm Age	Above Median	3,555	Real Estate Agents	Residential	337
	Below Median	2,844		Commercial	159

Notes: The counts are numbers of recordings. The groups are those shown in Figures 5 and 6.

References

Blomgren, Michael, Yang Chen, Manwa L. Ng, and Harvey R. Gilbert. "Acoustic, aerodynamic, physiologic, and perceptual properties of modal and vocal fry registers." *Journal of the Acoustical Society of America* 103(5), (1998): 2649–2658.

Boersma, Paul. "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 (1993)*: 97–110.

Boersma, Paul. "Praat, a system for doing phonetics by computer." *Glot International* 5, no. 9 (2001): 341–345.

Wolk, Lesley, Nassima B. Abdelli-Beruh, and Dianne Slavin. "Habitual use of vocal fry in young adult female speakers." *Journal of Voice* 26(3), (2012): e111–e116.