Gender identity and mode-switching behavior: Evidence from the human voice*

Yosh Halberstam

November 2026

Abstract

Studies of the human voice typically summarize a speaker by average pitch. Using about 40,000 voicemail greetings of lawyers at top U.S. law firms, a male-dominated work environment, I show that about a third of women alternate between their primary pitch near 200 Hz and a secondary mode near 100 Hz, where men's voices sit. Men show no second mode, and the pattern is robust to the pitch-tracking settings tested. Mode-switching is more common among junior women, and listeners perceive women who mode-switch as lower ranking. Auxiliary data find the pattern in other workplace settings. The pattern is consistent with workers moving between competing norms, often called code-switching, and shows what average-based measures of behavior can miss.

1 Introduction

Accounts of women under pressure to lower their voices at work are easy to find.[1] In one NPR story, a New York litigator recalled that the only feedback a partner gave her on a presentation was that her voice was "very high," so she set out to lower it (Starecheski, 2014). She is not alone: voice coaches report that women in finance hire them to bring their pitch down so they are taken seriously (Peters, 2022). The best-known recent case is Elizabeth Holmes, who spoke in a deep voice while running Theranos. She later described it as part of a persona she created to be "taken seriously," and she has since dropped it (Chozick, 2023).[2]

Voice is one part of how people communicate at work, and employers value communication highly. It ranked first on LinkedIn's list of the most in-demand skills for 2024, ahead of every technical skill (Royle, 2024).[3] LinkedIn's Ryan Roslansky and Aneesh Raman argue that skills like communication

*I thank numerous colleagues for very helpful suggestions as well as seminar and conference participants at Bar-Ilan University, the CFPB, Colgate University, Columbia University, the FDIC, the Institute for Defense Analyses (IDA), Queen's University, York University, the University of Rochester, the University of Toronto, the NBER Summer Institute joint workshop in Labor and Personnel Economics, and the CIFAR IOG program. Joyce Chen, Raheeb Dastagir, Katherine Lavrova, David Rhodes, Sida Shuai, Jingyuan Wang, Fangyu Yan, Jinguang Yang, Ruiqi Yu and Austin Zheng provided outstanding research assistance. This work was partially supported by the Social Sciences and Humanities Research Council of Canada. Email: yosh.halberstam@gmail.com.

[1] Growing attention in media (e.g., www.npr.org/sections/codeswitch/, www.girlboss.com/identity/code-switching-at-work, and www.linkedin.com/news/story/the-cost-of-code-switching-5434538/) and in arts (e.g., Boots Riley's film "Sorry to Bother You").

[2] An earlier example is Margaret Thatcher, who is widely reported to have trained with a voice coach to lower her pitch on her way to becoming prime minister.

[3] An OECD study finds that demand for social and language skills grew most in the occupations most exposed to AI (Green, 2024, using Canadian job postings).

will matter more as AI takes over routine and technical work (Roslansky and Raman, 2026), and Nobel laureate Philippe Aghion has called for rewarding firms that build soft skills such as working with others (Aghion, 2025). Firms already sell AI tools that change how workers sound, converting their accents in real time (Wiggers, 2025; Fisher, 2026).

Yet how workers adjust the way they sound is hard to see. Accounts of it are mostly anecdotal, and economic research has focused on hard skills that are easier to measure (Deming, 2017). This raises a measurement question: how do workers adapt how they sound to fit expectations at work?

Conforming to workplace norms is costlier for some workers than others. Because norms are often set by the in-group, the out-group has to change more to follow them. Culture, norms and identity shape well-being, productivity and group dynamics at work, but theory has run ahead of evidence. The reason is measurement. Workplace communication is confidential and sensitive, and much of it is nonverbal and subtle, so outside of experiments it is rarely observed.

Studies of discrimination against the out-group typically focus on fixed attributes of workers, such as sex and race. Yet, malleable worker characteristics, such as voice, may reflect, rather than determine, outcomes in the labor market. Indeed, pressures to conform to market norms may influence identity choices (Akerlof and Kranton, 2000). Accounts from workers, especially African American and female workers in the U.S., describe a behavioral response to these pressures that moves between in-group and out-group norms rather than adopting either one.

Voice is one place to look. It shapes how people are judged (Grogger, 2019), and people can change how they sound. In a male-dominated workplace, there may be pressure to sound authoritative, like the men. Full conformity would mean adopting a fixed mode at work, the way someone raised in a non-English-speaking country speaks their native language at home and English at work. If a subgroup of women adopted a man's pitch at work, as Holmes did, average pitch across women would show a separate low cluster. It doesn't: across women at elite U.S. law firms, average pitch forms a single bell curve near 200 Hz (Figure 8).

The average hides it. In this large-scale study of the human voice, recordings of workers in a male-dominated work environment show that about one third of women subtly alternate between two distinct frequencies, about 100 Hz and 200 Hz, within a brief greeting (two-mode women; the rest are one-mode). Mode-switching is more common among junior women, and auxiliary data I collected show it in other workplaces too. Listeners rate one-mode women as more dominant and senior than two-mode women. Men show no second mode, and the pattern is robust to the pitch-tracking settings tested. The contribution is measurement: a behavior that averages hide, documented at scale. Its causes and costs are the next questions.

Unlike other examples of nonverbal behavior, such as how individuals dress or present themselves at work, voice frequency can be measured precisely and at scale, and it is less directly tied to income than dress. What has changed since this project began is how cheap such measurement has become: with AI tools, the same approach can now be applied to calls, meetings and other recordings of work.

2 Background and Preliminary Findings

Research on how the brain processes vocal cues has advanced substantially (Scott, 2019). fMRI analysis has enabled significant scientific progress in understanding how the human brain distills meaning from sound (e.g., McGettigan and Scott, 2012; Formisano et al., 2008; Mathias and von Kriegstein, 2014; Creel and Bregman, 2011; Weston et al., 2015). This research provides a framework for a large literature in the social sciences on snap judgements (Creel and Bregman, 2011; Weston et al., 2015). That literature links vocal cues to how listeners perceive a speaker's attributes (Imhof, 2010; Baus et al., 2019; Grogger, 2011; Chen et al., 2016; Burgoon et al., 1996; O'Hair and Cody, 1987). Measurement of how people produce their voices has moved more slowly.

Until recently, studies of human voice used small samples (e.g., Leongómez et al., 2017; Pisanski et al., 2016; Smith and Patterson, 2005; Banse and Scherer, 1996), which made distributional patterns in human speech hard to detect. Large-scale measurement has since become common, helped by automated speech tools: recent studies measure pitch and vocal tone across tens of thousands of earnings calls and parliamentary speeches (Rittmann, 2024; Baik et al., 2025; Ewertz et al., 2026; Rask and Hjorth, 2026).

Human anatomy enables one to manipulate the vocal cords finely and rapidly. However, existing studies have largely focused on a person's *average* pitch (Klofstad et al., 2012; Apple et al., 1979; Tigue et al., 2012; Ekman et al., 1976; Mayew et al., 2013), leaving the distribution of pitch within a speaker unexamined. This study shows that the full distribution of pitch can reveal a richer set of vocal phenomena. In the context of the male-dominated law industry on which I focus, a significant proportion of female lawyers alternate between a primary female mode at about 200 Hz and a secondary mode at about 100 Hz that is coextensive with the primary (and only) male voice frequency mode.

Table 1. Firm descriptive statistics. This table shows summary statistics for the final 84 firms used in the main dataset. The data were collected from several external sources. Firm rank comes from Vault.com, where 76 firms were among the top 100 in 2017, and 8 firms were among the top 100 in adjacent years: 3 in 2018, 3 in 2016, and 2 firms appeared in both years (but not in 2017). Productivity measures come from the 2018 Am Law 200, data on lawyer counts and gender composition come from the 2018 NLJ 500 and 2018 NLJ Female Scorecard. Year established for merged firms is based on oldest firm at the time of merger. Several firms do not disclose their data and do not rank among the top revenue generating law firms.

Variable	Obs.	Mean	Std. Dev.	Min	Max
Firm Rank	84	54.67	30.40	3	100
Share female	77	0.36	0.03	0.25	0.44
Share partners female	77	0.21	0.03	0.13	0.28
Total lawyers	80	940.96	785.76	106	4719
Lawyers per office	80	62.98	41.64	21.33	331
Revenue rank	78	54.63	36.34	1	155
Total revenue (billions US$)	78	0.92	0.68	0.18	3.16
Profit per partner (millions US$)	78	1.30	0.98	0.20	4.55
Year established	84	1918.58	46.26	1792	2013

To begin, I collected a large sample of voicemail greetings of workers. The main sample comes from lawyers at top private law firms in the United States. The Vault 100 firms that I study accounted for about 25 percent of total revenues in the legal services industry at the time.[4] Firm-level descriptive statistics gathered from external sources of data about law firms are presented in Table 1.[5] The average number of lawyers per firm is 941. The average profit per partner is $1.3 million. The oldest firm in the dataset was established in 1792 and the youngest in 2013. Although these firms vary along several dimensions, they are extremely homogeneous with respect to female representation. On average, 36 percent of lawyers within a firm are female. Among partners, 21 percent are female. The standard deviation of each of these two measures is 3 percent. This imbalance is typical of other high-skill professions and corporate roles in the U.S., where females are significantly underrepresented in top positions.

To build the dataset, I scraped the phone directories from each firm's webpage, used call-management software to call each phone, and recorded the voicemail greeting once the call was connected. Most calls were made in late December 2017 and early January 2018, after pilot calls starting in May 2017, primarily during weekend nights to maximize the chances of reaching the lawyers' voicemails. Each

[4]The Vault 100 is a ranking generated from survey responses of approximately 20 thousand associate lawyers each year and is highly correlated with firm revenues. Based on the Census NAICS (North American Industry Classification System) Code 5411 ("legal services"), total revenue in this industry is approximately 1/3 trillion dollars (2019 Quarterly Services Survey) generated by over 1.1 million employees across 175 thousand law offices in the US (2016 County Business Patterns).

[5]Based on a pilot study, I dropped firms that either had a live receptionist 24/7 (3 firms) or firms that had less than 10 percent of voicemail greetings self-recorded by the lawyer.

Table 2. Lawyers by title and gender. This table presents the number of recordings in the main dataset of lawyer voicemail greetings by lawyer gender and job title.

	Associate		Counsel/Other		Partner		All	
Gender	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Female	7,435	44.67	2,407	38.96	3,902	22.77	13,744	34.39
Male	9,209	55.33	3,771	61.04	13,238	77.23	26,218	65.61
Total	16,644	100	6,178	100	17,140	100	39,962	100

Table 3. Likelihood that a voicemail greeting is self-recorded. This table summarizes the distribution of voicemail greetings in the main dataset by lawyer gender and the probability that the greeting was self-recorded by the lawyer. See Section C for details on the machine learning classification for the probability that a voicemail greeting was self-recorded by the lawyer.

	Prob > 0.95		Prob > 0.50		Prob < 0.05		All	
Gender	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Female	3,711	25.83	7,545	35.25	3,551	28.15	13,744	34.39
Male	10,654	74.17	13,858	64.75	9,065	71.85	26,218	65.61
Total	14,365	100	21,403	100	12,616	100	39,962	100

of the recordings I obtained was then trimmed to contain only the first 3 seconds. This timeframe minimizes the likelihood of capturing silence or machine-generated audio, such as generic instructions for leaving a message. My final sample comprises 39,962 lawyers across 690 offices employed by 84 law firms that were listed in the annual Vault 100 prestige rankings between 2016 and 2018 at least once. For these lawyers, I merged demographic information obtained from the ALM Legal Compass database, a leading directory of lawyers, into the dataset by phone number and lawyer name. Table 2 summarizes these data by title and gender. The share of female lawyers in the data is consistent with the externally obtained firm-level data in Table 1 referenced above. Most striking is the difference between female representation at the Associate level (45 percent) relative to the Partner level (23 percent).

Standard quality digital recordings provide one data point per 1/8 of a millisecond of playback time, each representing the approximate amplitude at that specific moment. These samples comprise the raw digitized audio data. To estimate the audio frequency at a given point in time, I use a 60-millisecond analysis window containing 480 amplitude data points split evenly on either side of the estimation point. This window length corresponds to three cycles of a 50 Hz signal, the minimum detectable frequency I set for this analysis and well below the range of frequencies in the natural voice register. To account for the local nature of estimating the frequency, amplitude data closer to the estimation point receive more weight than those farther out in the analysis window. See Section B for technical details and robustness checks.

The frequency is defined as the inverse of the time (in seconds) it takes for a soundwave to repeat itself. As is standard in studies of the human voice, the frequency used in this study is the fundamental frequency, which is the lowest frequency of a periodic waveform. Mersenne's law is often used to model the connection between the frequency and one's vocal cord properties, tension (T), tissue mass per unit length (μ) and length (L):

$$F_0 = \frac{1}{2L}\sqrt{\frac{T}{\mu}} \tag{1}$$

The basic approach finds the candidate frequencies at which the signal best matches a shifted copy of itself (the highest autocorrelation). This mechanically favors lower frequencies: a signal that repeats every 5 milliseconds also repeats every 10, 15 and 20 milliseconds, so the method adjusts for it.

The first step is computing the autocorrelations of the analysis window for each 1/8 of a millisecond lag (the sampling rate) until a maximum lag of 20 milliseconds. This corresponds to the range of 50 to 4000 Hz (the Nyquist frequency). Areas where the autocorrelation values switch from increasing to decreasing are identified as local maxima, and the corresponding frequencies are used as frequency candidates.

Using 5-millisecond time steps, where the analysis window is shifted 5 milliseconds (or 40 data points) at a time, three seconds of playback translate into 589 frequency estimates, where the first and last estimates are given at points 0.03 and 2.97 seconds, respectively. To select the most likely frequency estimate at each point in time, I impose a post-estimation ceiling of 400 Hz, which is well above the range of human voice frequencies produced by the natural voice register, and choose the candidate (if any) associated with the highest strength, subject to exceeding a minimal strength threshold. Because the 3-second clips contain periods of silence and noise, the actual number of frequency estimates per clip is significantly lower than 589 and varies from clip to clip.

Finally, not all workers record their own voicemail greeting. I combine machine learning techniques with biographical information about the lawyers as well as verbal and nonverbal characteristics of the greeting to separate self-recorded greetings from greetings in third person made by assistants or generic automated call management operators. Each recording is assigned a number from 0 to 1 representing the likelihood that the voicemail greeting was personally recorded by the lawyer. Table 3 summarizes the results of this classification by lawyer gender. Approximately one-half of the voicemail greetings are classified as self-recorded with a probability greater than 0.5, whereas about one-third of the recordings are assigned a likelihood greater than 0.95.

There are several challenges to visualizing the frequency data. First, there are millions of frequency estimates and plotting them all would result in an incomprehensible figure. Second and more substantively, audio clips with minimal noise and periods of silence result in many frequency estimates but others result in a significantly smaller number of estimates. Not accounting for these differences would overweight the former clips relative to the latter in the figure. Further, as mentioned above, not all greetings are self-recorded. To address these issues, for each of the 589 points in time, I scatter frequency estimates from 1,000 lawyers with the highest probability of self-recorded greetings. Because periods of silence and noise vary from recording to recording, the clips from which the estimates come vary from point to point.

Figure 1 shows this subsample of frequency estimates for female and male lawyers separately. The horizontal axis in the figure represents time elapsed from the beginning of the recording, whereby frequency estimates are given every 5 milliseconds of playback time. As expected, most estimates for females are significantly higher than those for males. However, a band of significantly lower estimates for females (approx. 100 Hz) is also apparent. Figure S5 shows two individual greetings with the same average pitch, one of which moves into this band. This is the key finding that motivates the analysis in the study.

3 Between- or Within-Person Bimodality?

Does this secondary mode reflect a group of women who consistently speak like men, or individual women alternating between two modes (or both)? I undertake the following steps to answer this question. First, to address the variability in the total number of frequency estimates obtained from each recording, precisely 100 estimates are selected from each clip. Specifically, each of the selected estimates corresponds to a percentile, thereby retaining acoustic information necessary to consistently represent the shape of each clip's density. In contrast, random sampling is unsuitable because it tends towards a unimodal or uniform density, thereby obscuring the true shape of each clip's density.

Let $h_i(\,)$ be the empirical voice frequency density function for clip i. Then, for $p = 1, \ldots, 100$, each lawyer's voice frequency percentile in Hz, $F_0^{i,p}$, is a number defined as the highest frequency estimate of clip i such that:

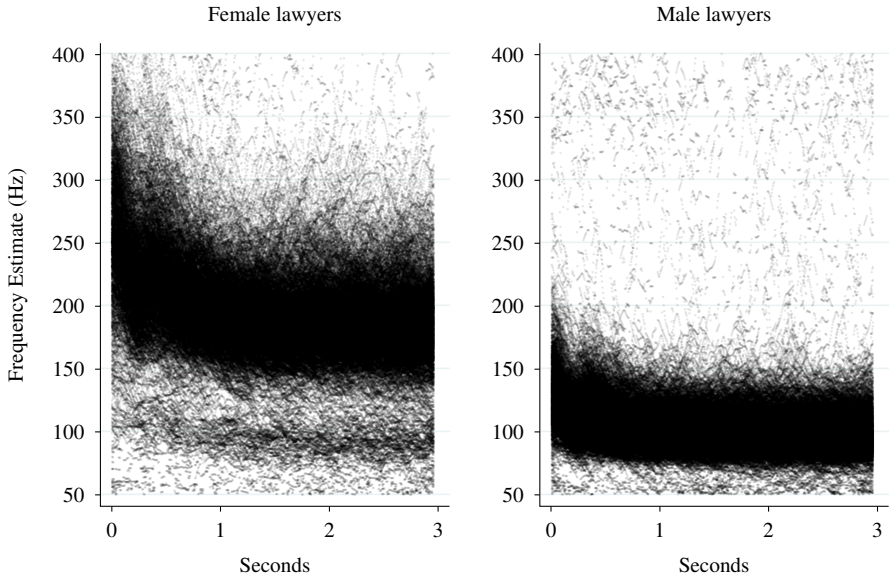

Figure 1. Frequency estimates over time, female and male lawyers. For each 5-millisecond point in time running from 0.03 to 2.97 seconds, there are 1000 frequency estimates scattered for females (left) and males (right), respectively, from recordings assigned the highest likelihood of containing a self-recorded greeting.

$$\int^{F_0^{i,p}} h_i(\nu)\,\mathrm{d}\nu \leq p/100 \tag{2}$$

Second, the mean frequency estimate of each clip is subtracted from each of the individual percentile estimates ($F_0^{i,p} - \overline{F}_0^i$). These demeaned estimates neutralize level differences in pitch between voicemail greetings.

Figure 2 presents results of kernel density estimations and histograms using these data. To directly compare females to males, the density of each group is shifted by the group's mean frequency estimate. Using the probability assigned by the machine learning model, the subfigure on the right uses data from clips of 21,403 voicemail greetings that *more likely than not* were recorded in first person by the lawyer, whereas the subfigure on the left uses data from a subset of these clips where this likelihood exceeds 95 percent (n = 14,365). The latter figure provides a more accurate representation of the density albeit at the cost of using a smaller sample. In both subfigures, the female density has a secondary mode around 100 Hz, which is where the modal male voice frequency lies, a phenomenon I call *mode-switching*. In contrast, no such secondary modes are detected among the male densities. These figures indicate that the secondary mode is not entirely driven by differences between voicemail greetings or by differences in the number of frequency estimates per greeting.

A common test for multimodality in the statistics literature draws from Silverman (1981), who provides a theoretical framework to assess whether the true unobservable population density has a specific number of modes. The basic idea builds on the property that when using a Gaussian kernel to estimate a density, the number of modes is a decreasing function of the bandwidth. In plain terms, the more you smooth a histogram, the fewer peaks it shows. The standard method entails (1) locating for every m = 1, 2, ... the smallest bandwidth ('critical bandwidth') that can support m modes or less, (2) generating smoothed bootstrapped samples from each critical density using a Gaussian kernel, and (3) estimating the density of each smoothed bootstrapped sample using the critical bandwidth. The proportion of samples with greater than m modes reflects the significance (i.e., p-value) of the critical bandwidth. A low p-value is evidence against the null hypothesis that the underlying density has m or fewer modes. Put differently, if all the samples indicate m or fewer modes, then the kernel density must be significantly oversmoothed to remove the appearance of mode m + 1. The test is seen

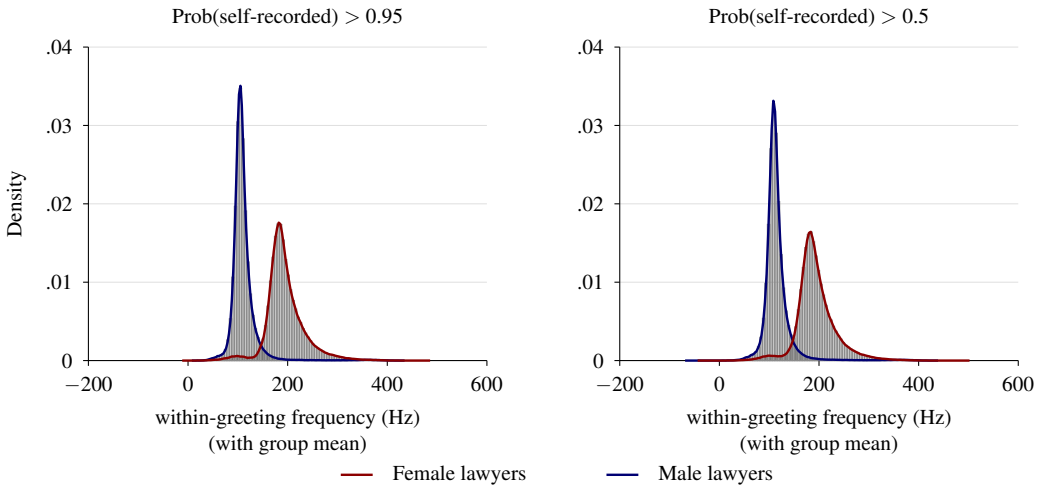

Figure 2. Demeaned frequency densities, female and male lawyers. To remove the influence of level differences between lawyers, *recording-level* demeaned frequency estimates are used to estimate the density of the voice frequency. Results are shifted rightward by the overall average frequency estimate of each gender group. The left graph uses recordings with a high (95%) threshold of the probability that the greeting was self-recorded by the lawyer, and the right graph uses recordings with a low (50%) threshold. Kernel densities (in solid lines) and histograms with 5 Hz bin widths (in gray bins) are indistinguishable.

as conservative since the bootstrapped samples are drawn from the critical density only and tends to underestimate the true number of modes.

Table 4 shows the results of the Silverman test against the null hypothesis that the underlying frequency density has m or fewer modes. For computational purposes, I used demeaned data from 1,000 randomly selected clips that were self-recorded by female lawyers. For each additional mode, the test is performed by creating 50 perturbed samples of the 100,000 frequency percentiles using the critical bandwidth identified with precision of 0.01 Hz or lower. The mode location refers to the distance from one's mean frequency, estimated using 40 averaged shifted histograms (ASH-WARPing). The results strongly reject the null hypothesis that the frequency density is unimodal ($p < 0.02$).[6]

Table 4. Silverman test for the number of modes. The table shows the results of a Silverman test for the null hypothesis that the underlying frequency density has m or fewer modes. This test used demeaned data from 1,000 randomly selected clips from the main dataset that were self-recorded by female lawyers. For each additional mode, the test is performed by creating 50 perturbed samples of the 100,000 frequency percentiles using the critical bandwidth identified with precision of 0.01 Hz or lower. The p-value is the proportion of samples that produced more modes than stated by the null. The mode location refers to the distance from one's mean frequency, estimated using 40 averaged shifted histograms (ASH-WARPing).

H_0	Critical Bandwidth (Hz)	Change (%)	Silverman p-value	Mode Location (Hz)
$m \leq 1$	9.61	–	0	−11.41
$m \leq 2$	3.6	63	0.98	−86.15
$m \leq 3$	3.59	0	0.96	235.97
$m \leq 4$	3.58	0	0.62	169.33
$m \leq 5$	3.46	3	0.48	−102.39

I next investigate how prevalent this phenomenon is among female lawyers in my sample.

[6]The test rejects a single mode but cannot reject two or more modes; however, modes 3 and above are more than 100 Hz from the mean and likely spurious rather than reflect robust features of the underlying density.

4 Is Mode-Switching Widespread?

About a third of the women turn out to have bimodal pitch. My empirical approach to finding them involves two steps: first, estimating the location of a low frequency mode in each individual clip. Second, classifying clips into groups based on the estimates from step one. In both steps, I use finite mixture models (FMM), a methodology extensively used to classify observations into groups (Deb, 2012; Deb and Trivedi, 1997).

To estimate the frequency modes in each recording, an iterative procedure flexibly searches for the best fit between the 100 percentiles described above and a mixture of normal distributions. Specifically, I approximate the density of each clip i using the following model:

$$h_i = \sum_{k=1}^{g} \pi_{i,k} N\left(\mu_{i,k}, \sigma_{i,k}^2\right) \tag{3}$$

In this model, g is a predetermined number of components in the mixture, and $\mu_{i,k}$, $\sigma_{i,k}^2$ and $\pi_{i,k}$ are the component-specific mean, variance, and the share of component k ($\sum_k \pi_{i,k} = 1$) respectively, to be estimated from the percentile data.

To increase precision, each clip was screened by (at least) one human listener. I focus on the sample of recordings classified as self-recorded by female lawyers. This process yielded 6,399 voicemail greetings. The *mean* voice frequency in this sample is 195 Hz. Some experimentation indicated that five components ($g = 5$) were optimal to fit the individual densities and detect the secondary mode. Table 5 shows aggregate model fit for a variety of mixture models using the 100 demeaned percentiles of a random subsample of 1,000 clips from all 6,399 clips in the sample of verified self-recorded female lawyers. In general, lower values of Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC) are better, but computational costs grow exponentially with g, the number of components in the mixture. The rightmost column of the table shows that $g = 5$ maximizes model fit according to BIC, and the marginal improvement from $g = 5$ onwards is below 1% according to AIC.

Table 5. Fit of finite mixture models by number of components. This table shows aggregate model fit for a variety of mixture models using 100 demeaned percentiles of a random sample of 1,000 clips from all 6,399 clips in the sample of verified self-recorded female lawyers. The results show that $g = 5$ maximizes model fit according to Bayesian information criterion (BIC), and the marginal improvement from $g = 5$ onwards is below 1% according to Akaike's information criterion (AIC).

Model	Obs.	ll(model)	df	AIC	AIC (% change)	BIC	BIC (% change)
fmm1	100,000	−509800	2	1019604		1019623	
fmm2	100,000	−497941	5	995891.9	2.3256	995939.5	2.3228
fmm3	100,000	−496770	8	993556.3	0.2345	993632.4	0.2317
fmm4	100,000	−496509	11	993040	0.0520	993144.7	0.0491
fmm5	100,000	−496462	14	992952.6	0.0088	993085.8	0.0059
fmm6	100,000	−496445	17	992924.6	0.0028	993086.4	−0.0001
fmm7	100,000	−496441	20	992922.5	0.0002	993112.8	−0.0027

In Figure 3, I present clip-level estimation results of the lowest frequency mode location (i.e., $\min_k \mu_{i,k}$). The histogram shows two distinct clusters of estimates consistent with two types of frequency densities. To group the estimates, I use a two-component mixture model ($g = 2$). The predicted density based on this model is depicted in the same figure with a dashed line along with the predicted individual normal distributions in the mixture. The implied cut-off of 115 Hz between the two groups puts 36 percent of female lawyers in the two-mode group (delta method standard error 0.7). Their low mode estimates sit more than 80 Hz below their *average* frequency estimate.

For robustness, Table 6 shows how the estimated shares of two-mode (bimodal) and one-mode (unimodal) clips vary with alternative specifications for estimating frequency modes. Rows 1-6 use a

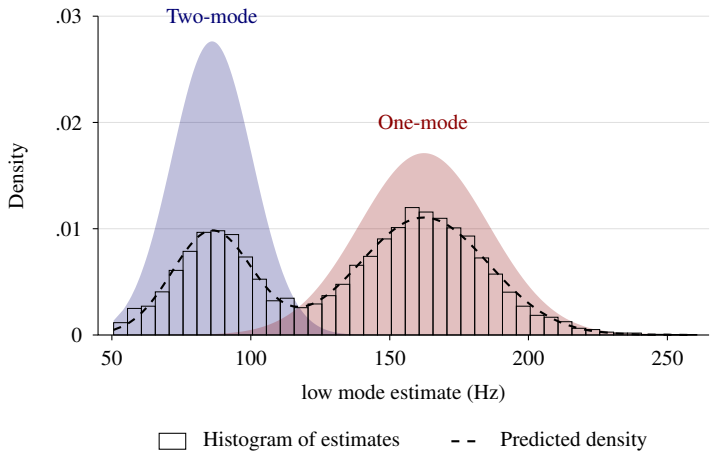

Figure 3. Estimated location of each greeting's lowest mode, female lawyers. The location of a low mode is individually estimated for each of 6,399 verified self-recorded greetings of female lawyers. The estimation results indicate two clusters of estimates defining two groups, where the two-mode group comprises 36% of the recordings. The smooth curves are the individual FMM components.

random subsample of 1,000 clips from these 6,399 clips, though the exact number of observations may vary slightly due to failure of FMM convergence. Row 7 uses the full sample of 6,399 clips, and uses gamma instead of normal densities with a 5-component mixture. The estimated share of bimodal clips stabilizes from Row 4 ($g = 5$) onwards, supporting Table 5 in showing that $g = 5$ maximizes the model fit.

One concern is a known quirk of pitch tracking. A sound wave that repeats every 5 milliseconds (200 Hz) also repeats every 10 milliseconds, so an algorithm can report half the true pitch (100 Hz). The algorithm used here guards against this by favoring the shorter period (the higher pitch) when the two fit about equally well. Strengthening that preference, or changing the analysis window, leaves the share of two-mode women between 33 and 37 percent. Switching off a second safeguard, Praat's penalty on sharp changes in pitch, raises the share to about two-thirds (Table S2), which shows how much the estimate depends on this protection. The low band also appears throughout the greeting and sits near 100 Hz, well above the roughly 50 Hz typical of vocal fry (Blomgren et al., 1998; Wolk et al., 2012; Figure S6).

Table 6. Robustness of bimodality to low mode measures. This table shows how the estimated shares of two-mode (bimodal) and one-mode (unimodal) clips vary with alternative specifications for estimating the clip-level frequency modes. Results are from a 2-component FMM applied to the different individual low mode estimates. The number of observations column refers to the clip-level FMM estimation of the random sample of 1,000 clips from the 6,399 verified self-recorded greetings by female lawyers. Missing observations are due to clip-level convergence failures. The bottom row presents results from the full sample of verified female lawyers, substituting normal with gamma densities in a 5-component mixture. The results are similar for all but the 2 and 3 component models. For these, the mode shares were significantly skewed toward the primary mode relative to the other results.

Row	Individual FMM Components	Group	95% C.I. (Share)	95% C.I. (Location)	Group	95% C.I. (Location)	Obs.
1	2	Two-mode	(0.085, 0.124)	(84.306, 90.325)	One-mode	(179.291, 182.219)	1,000
2	3	Two-mode	(0.275, 0.335)	(84.854, 88.660)	One-mode	(169.137, 172.829)	999
3	4	Two-mode	(0.324, 0.389)	(84.979, 88.612)	One-mode	(163.034, 167.051)	998
4	5	Two-mode	(0.348, 0.416)	(84.535, 88.411)	One-mode	(160.970, 165.396)	1,000
5	6	Two-mode	(0.371, 0.439)	(84.177, 87.852)	One-mode	(159.477, 163.799)	990
6	7	Two-mode	(0.365, 0.436)	(84.496, 88.500)	One-mode	(157.795, 162.334)	989
7	5 (Gamma)	Two-mode	(0.371, 0.398)	(87.362, 89.116)	One-mode	(163.494, 165.054)	6,399

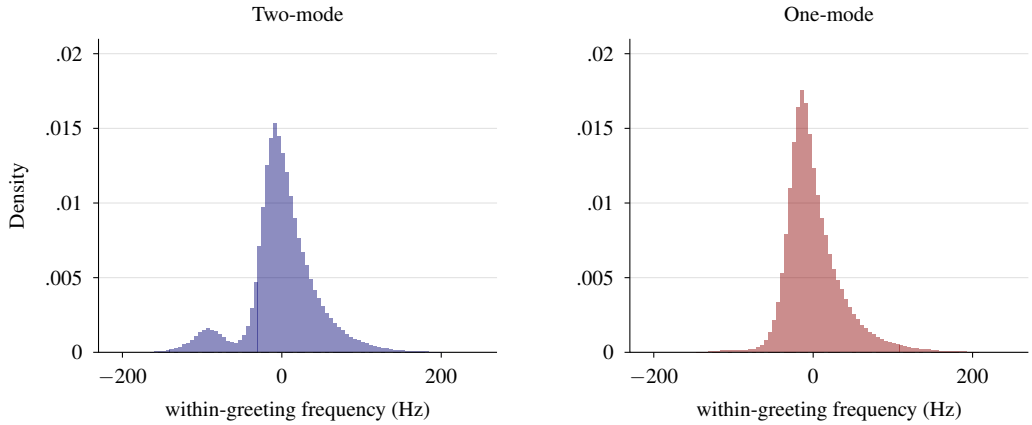

Figure 4. Demeaned frequency histograms by mode group, female lawyers. Based on the two-mode and one-mode classification in Figure 3, histograms of the demeaned frequency estimates are shown for each group of female lawyers. The results show a unimodal density for the one-mode group and a bimodal density for the two-mode group with a low secondary mode accounting for 8.5% (standard error 0.20 percentage points) of the density.

Using the demeaned percentiles described above, I plot the histograms of each group of female lawyers in Figure 4. The histogram of female lawyers associated with the high values of low mode estimates (one-mode group) shows no sign of bimodality, whereas the histogram of the two-mode group clearly displays a bimodal density. For this group, the primary mode is located at 197 Hz and the secondary mode is located at 96 Hz, where the latter accounts for 8.5 percent (delta method standard error 0.2 percentage points, adjusted for clip-level clustering) of the 5-component mixture density.

5 Is Mode-Switching Context-Specific?

Does the pattern vary? It is more common among junior women, and data from several other sources show it in other workplace settings too. Using the same mixture model methodology described earlier, I summarize the estimation results in Figures 5 and 6. In these figures, I show the estimated fraction of females with bimodal speech including 95 percent confidence intervals. In each row, the total number of recordings used to estimate the mixture model is reported in Table S4.

In Figure 5, I present results using the main sample of 6,399 female lawyers. Starting at the top of the figure, I show estimation results for lawyers who include litigation as one of their practice areas versus those who do not. I hypothesize that litigators may interact differently with clients and judges relative to non-litigators; however, I find no difference between the two groups. In contrast, the incidence of bimodal voice patterns does significantly vary by seniority. Voice frequency densities of 31 percent of senior lawyers, including Partners and Counsels, are classified as bimodal, yet 43 percent of all Associates mode-switch. Many factors could drive this difference. Observable characteristics explain little of the pattern itself: removing graduation-year, firm, title and litigator effects barely changes the low-mode estimates (correlation 0.985), and classifying the residuals yields a two-mode share of about 36 percent, the same as at baseline (Table S2).

The subsequent categories in the figure focus on differences between firms. In general, I find no evidence for variation in the incidence of mode-switching based on firm prestige, age or female representation. One reason for this could be that the firms that I study represent a very homogeneous sector. For example, based on 2018 ALM data (NLJ 500 and NLJ Female Scorecard), the average share of females in these firms was 0.36 with a standard deviation of 0.03 indicating limited between-firm variation (Table 1). Likewise, also with a standard deviation of 0.03, the average share of female

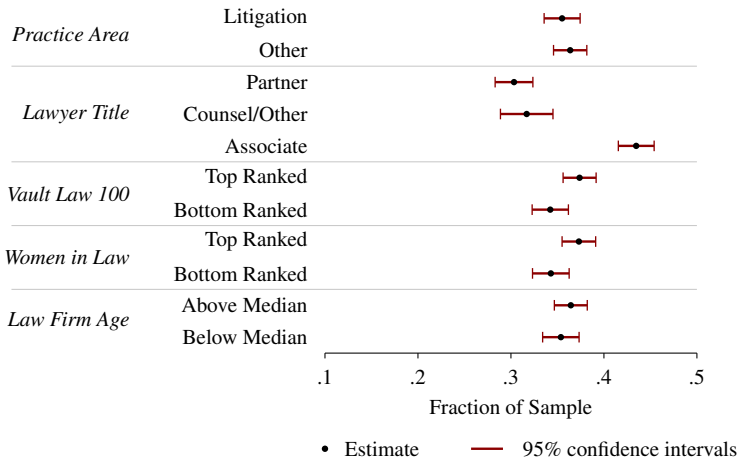

Figure 5. Share of bimodal recordings by lawyer characteristics, main sample. This figure displays the estimated share (with associated 95% confidence intervals) of recordings in each sample that contain bimodal frequency densities. The estimation procedure is based on a finite mixture model (FMM) methodology further explained in the text. Using the sample of 6,399 self-recorded voicemail greetings of female lawyers, the number of clips in each subsample is reported in Table S4. Six unranked firms were omitted from the Women in Law comparison. Vault Law 100 and Women in Law are firm rankings. See text for description of subsamples.

partners was 0.21. This homogeneity may be reflected in the distribution of behaviors in these firms, including speech patterns.

I next turn to estimation results using auxiliary data. These samples are significantly smaller in size than the main sample of female lawyers as reflected in the wider confidence intervals in Figure 6 relative to Figure 5. Nonetheless, these data are meant to address three related questions: First, is the relatively lower incidence of mode-switching among senior lawyers reflected prior to or only after a promotion? Second, does the tendency to mode-switch persist after switching jobs or beyond the first few seconds of an introductory sentence? And third, do similar findings emerge in other professions, specifically female-dominated ones?

To answer the first question, it would be ideal to compare the voicemail greetings of workers before to after they get promoted. However, it is rare for a worker to change their voicemail greeting immediately following a promotion (and in general). Instead, I analyze the set of Associates from the main sample that were promoted within about two years. The estimates indicate that approximately 40 percent of this subsample mode-switch, which is not significantly different from the estimated 43 percent of all Associates. This suggests that the lower incidence of mode-switching among female senior lawyers is not explained solely by selection (i.e., mode-switching prior to promotion), and is consistent with a change in behavior after promotion, though the data cannot show this directly.

One context that forces a worker to change their voicemail greeting is switching jobs. About two years after the initial data collection, I recruited MTurk workers to check the lawyers' webpages and follow up on broken links. I was able to analyze the voicemail greetings of 198 female lawyers at their new place of work. Overall, before and after the move alike, slightly more than one third of the greetings are bimodal (two-mode).[7]

To investigate persistence in mode-switching over the duration of a speech, I use data from oral arguments at the U.S. Supreme Court. In these arguments, the opening sentence of each lawyer is: "Mr. Chief Justice, may it please the Court" and the time allocated to each lawyer is 30 minutes. Although the set of lawyers who argue in the Court is highly specialized, this context allows me to examine lawyers outside their firm as well as whether my findings extend beyond introductory

[7]More details about this sample are available in Section A.4. At the individual level, 48 percent of lawyers classified as two-mode at their first firm were also two-mode at their new firm. The sample is too small for firm conclusions.

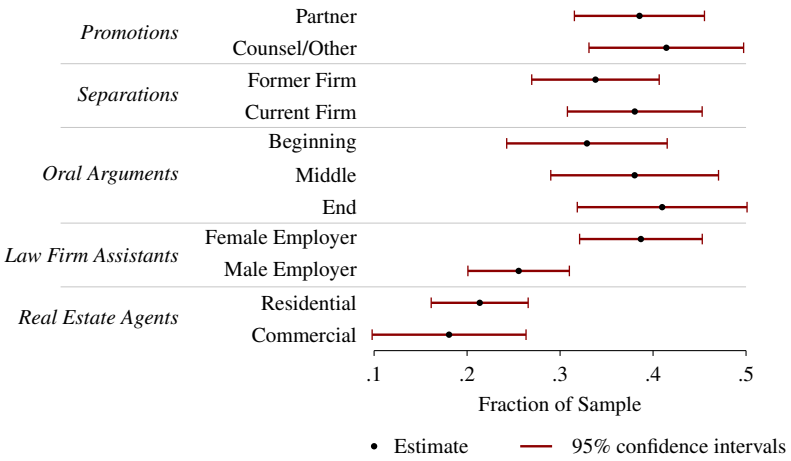

Figure 6. Share of bimodal recordings, auxiliary samples of females. This figure uses supplemental data. The number of clips in each subsample is reported in Table S4. The samples are promoted associates, lawyers who switched firms, Supreme Court oral arguments, law firm assistants and real estate agents. See text and Section A for more details about these samples.

sentences in an equally male-dominated environment (Biskupic et al., 2014). Data from 129 oral arguments made by female advocates between 1985 and 2005 suggest that they do.

Recordings of these arguments are publicly available. I collected three voice samples from every recording, each trimmed to 3 seconds. The samples are from the opening sentence, closing sentence, and one sentence taken from the middle of the argument (approximately minute 15). Using the opening sentence data, I find similar results to those of senior lawyers in the main sample: 33 percent of the advocates were classified as bimodal. Beyond the first 3 seconds, the estimates suggest that 38 percent of the middle-of-argument sample and 41 percent of the end-of-argument sample have a secondary mode; the differences are statistically insignificant. Overall, the findings identify mode-switching as a phenomenon outside the office and beyond the introductory sentence.

The next set of results comes from two female-dominated professions: executive assistants, and real estate agents. Beginning with the former, I analyzed voicemail greetings recorded by female executive assistants on behalf of a lawyer. Given the salience of gender identity in this article, I estimate the mixture model for assistants employed by male and female lawyers, separately. I find that assistants employed by female lawyers mode-switch at a rate of 39 percent. However, assistants employed by male lawyers are significantly less likely to mode-switch: only 26 percent are classified as bimodal.

Finally, I analyze data from RE/MAX, a large American real estate franchise. Like lawyers, real estate agents must be licensed to practice. Females comprise a large majority of the sector (58.9 percent of 1.1 million workers based on the 2019 Current Population Survey). Overall, I find the lowest incidence of mode-switching among this group of female workers. Bimodal vocalization is detected in voicemail greetings of only 21 percent of residential agents and 18 percent of commercial agents.

In sum, the pattern appears beyond lawyers and beyond the first few seconds of speech. It also varies with seniority and occupation, which a pitch-tracking artifact would be unlikely to do.

6 Perceptions from Bimodal Vocalization

The findings above indicate that a significant proportion of female workers mode-switch. Do listeners notice? Previous studies using fMRI show how the human brain distinguishes between high and low

frequency signals and uses this information to discriminate between males and females. Specifically, high frequencies consistently evoke a greater degree of cortex activation (e.g., Weston et al., 2015). Although humans may detect subtle acoustic differences between unimodal and bimodal clips, this does not imply that these differences provide meaningful labor market-related cues.

In a separate survey, I recruited 100 female and 100 male workers on Amazon's Mechanical Turk (MTurk). The recruitment was based on first-come-first-served and was limited to U.S. residents who completed at least 10,000 tasks on the platform with an approval rating of 99 percent or more. Raters were paid a flat fee. The workers ranged from age 18 through 74 with the median age being 36.

I paired 200 clips of female lawyers from the two-mode group with 200 clips of female lawyers from the one-mode group, where each pair had nearly identical mean frequency (within 1 Hz of each other). The selection of the pairs, including the order in which they appeared both within and across pairs, was randomized across workers. Each participant received a random set of 10 paired clips to rate on a 7-point Likert scale.

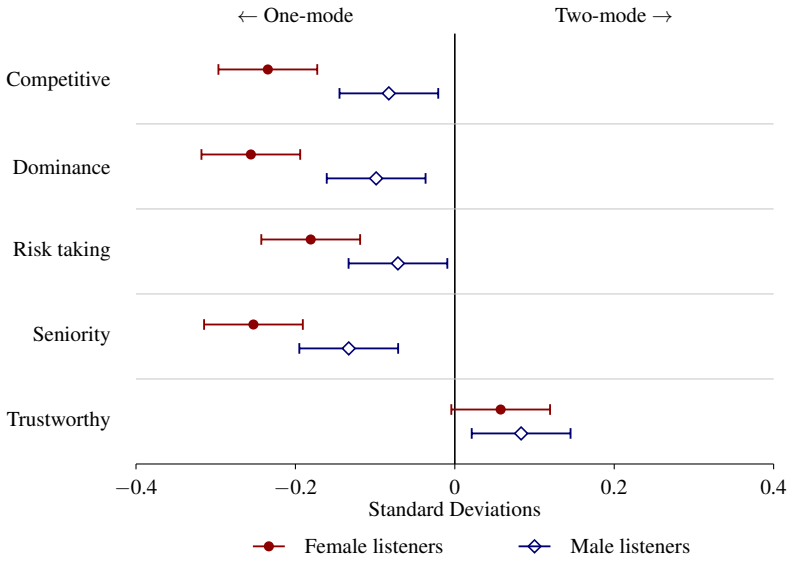

Figure 7. Listener ratings of two-mode relative to one-mode voices. 100 female and 100 male U.S. survey participants were recruited on Amazon's Mechanical Turk (MTurk). Each worker received 10 paired clips (<1 Hz difference in frequency means) to rate on a relative 7-point Likert scale. The distance between each point and the vertical line is the perceived difference between the two-mode and one-mode groups in terms of standard deviations of each attribute's ratings. Points to the left of the zero line mean that the one-mode voice was rated higher. Survey results suggest that humans, especially females, perceive unimodal vocalization significantly more dominant and high ranking than bimodal vocalization.

In the survey, each pair of clips was played and participants were asked to provide their relative impression of the lawyers on five attributes: competitiveness, dominance, risk-taking, seniority, and trustworthiness. Results from this survey are presented in Figure 7. Each point in this figure represents the mean deviation from a neutral rating (i.e., point 4 on the scale), scaled by the standard deviation of the attribute ratings. A point to the left of the vertical line means that workers perceived the attribute to resonate more strongly with lawyers from the one-mode group than the two-mode group. The results suggest a similar pattern for both male and female listeners: female lawyers with unimodal densities are perceived as more competitive, dominant, risk-taking and senior (but slightly less trustworthy) than female lawyers with bimodal densities. The differences perceived by females are significantly larger than by males. Female listeners rate unimodal vocalization about one quarter

of a standard deviation more dominant and senior than bimodal vocalization. For male listeners, the perceived difference is about half that size.[8]

In sum, results from this survey suggest that humans use the acoustic signal to inform their perceptions of the speaker.[9] One-mode lawyers are perceived as significantly more dominant and senior than two-mode lawyers.[10]

7 Mean Voice Frequency and Deep Voice Premium

Does a deeper voice pay off for women at these firms? To check, I describe the cross-sectional variation among female lawyers in the law firms that I study. For this exercise, I use the mean voice frequency estimate ($\overline{F}_0^i$) in each voicemail greeting.

As mentioned, a large literature studies the relationship between a person's mean voice frequency and other attributes of the speaker. Numerous studies have found that listeners tend to judge speakers with deeper voices more favorably. For example, speakers with deeper voices are perceived as more attractive, dominant, mature, and honest (Imhof, 2010; O'Hair and Cody, 1987). Other studies have found that they are perceived as more truthful and empathic, and to possess greater leadership capacity (Klofstad et al., 2012; Apple et al., 1979). One study by Mayew et al. (2013) uses data from quarterly conference call recordings of public companies listed in the S&P 1500 to find that CEOs (albeit all male) with deeper voices manage larger companies. Likewise, several lab experiments document volitional voice frequency modulation by speakers. For example, in a simulated interview, Leongómez et al. (2017) show that interviewees speak in a higher voice when randomly assigned to a higher status interviewer (e.g., by varying title). Relatedly, voice frequency has been shown to change concurrently with superficial exaggeration or reduction of body size by a speaker (Pisanski et al., 2016).

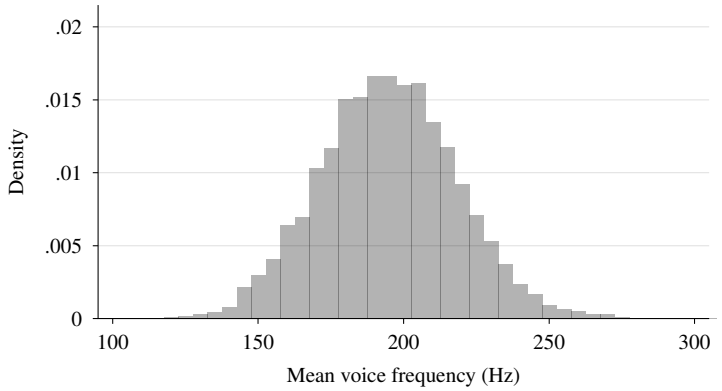

Figure 8. Distribution of mean voice frequency. This figure shows the histogram of mean frequencies for the 6,399 verified self-recorded female lawyer clips from the main dataset. The histogram shows a bell-shaped distribution around approximately 200 Hz, the primary female vocal mode.

[8] Pairs were matched on mean pitch, so a two-mode speaker's main voice sits slightly higher than her partner's. That difference does not drive the ratings. Holding fixed the share of time the two-mode speaker spends in the low mode, the gap between the two main voices has no detectable effect, while each additional 10 percentage points of low-mode time lowers ratings of dominance, seniority and competitiveness by about 0.2 to 0.3 standard deviations (the one-mode clip's main mode is proxied by its mean). The pattern holds for women and men listeners alike.

[9] Listeners rate the two groups differently even though they can hardly tell them apart when asked to pick the two-mode clip. In a separate test, listeners heard two-mode and one-mode clips with nearly identical mean pitch, played in reverse, and were asked which was two-mode. They were right 52.5 percent of the time, close to chance (Section E).

[10] For robustness, I explore the role of verbal content but do not find meaningful differences in the choice of words by both groups. See Table S3 for this comparison.

As a result, workers may choose to lower their voice to exploit these perceptual biases (Smith and Patterson, 2005). Particularly in the context of the male-dominated work environment that I study, females may wish to permanently adopt a "male" voice frequency. Figure 8 shows the histogram of the mean voice frequency for all 6,399 female lawyers in the main sample. As seen, the distribution is bell-shaped and centered around 200 Hz, the female vocal mode, with no sign of a group of women whose average pitch sits in the male range. This figure underscores how the use of the mean can be misleading, where the econometrician may conclude that the data generating process is unimodal.

Finally, I do not find strong evidence for a "deep voice premium" among female lawyers.[11] Although the mean voice frequency is strongly positively correlated with being an Associate (Figure 9, middle panel) relative to a Partner, the relationship becomes insignificant when controlling for years of experience (Figure 9, right panel), which is also a close proxy for age.

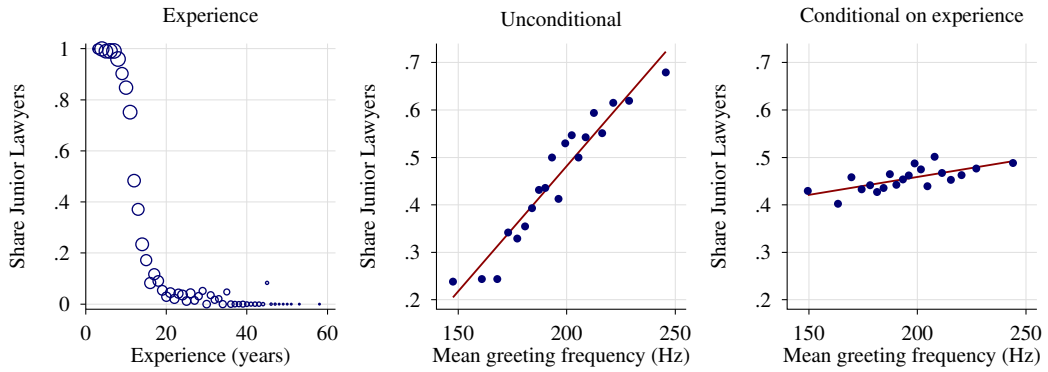

Figure 9. Experience, mean voice frequency and seniority among female lawyers. *Left panel.* This scatterplot shows the relationship between the share of Associates and experience (years from J.D.) among 4,682 female lawyers from the main dataset with experience data. The size of each circle is proportional to the number of female lawyers at each experience-year level. *Middle and right panels.* This figure shows binned scatterplots of an indicator for "Associate" and the mean frequency of 4,682 female lawyers from the main dataset with experience data. The middle panel shows a strong positive correlation between the mean frequency and the likelihood of being junior; however, as seen in the right panel, the relationship becomes significantly weaker when controlling for experience.

8 Conclusion

Workplace norms are often set by the in-group, so conforming to them asks more of the out-group (Akerlof and Kranton, 2000). In a male-dominated workplace, women may face more pressure than men to change how they behave and to "fit in," and their responses can be subtle.

Just as language can reflect identity (Auer, 1998; Myers-Scotton, 1995), so can nonverbal vocalization (Argyle, 1972). The mode-switching documented here suggests a parallel with code-switching, a concept that originated in linguistics (Gardner-Chloros, 2009; Heller, 1992) and has been extended to brief forms of out-group expression (e.g., Jeffries et al., 2015). Unlike other accommodative behaviors (Giles and Powesland, 1997), code-switching preserves the integrity of each underlying norm (Heller, 1988): a bilingual speaker who switches languages also switches accents. The women I study likewise move between two intact modes rather than blending them. Research links code-switching to psychological costs (McCluney et al., 2019).

[11] Relatedly, there is a body of research that specifically assesses a beauty premium in the market for lawyers. Using longitudinal data on graduates from one U.S. law school, Biddle and Hamermesh (1998) document a beauty premium with respect to lawyer earnings. They show that lawyers who meet normative beauty standards earn more than others. However, for the average worker, the returns to investing time or money to increase beauty have been found not to outweigh the costs (Lee and Ryu, 2012; Das and DeLoach, 2011), and an empirical challenge to exploring beauty-based code-switching is the difficulty of measuring beauty in an objective manner across time and space.

In economic models of identity, a worker typically conforms to the norms of one group or the other (Austen-Smith and Fryer, 2005), and market norms may reflect short-lived fashions (Pesendorfer, 1995) or persistent social customs (Akerlof, 1980). Social influence also shapes the choices of out-group members in schooling (Fryer and Torelli, 2010) and professional identity (Bursztyn et al., 2017), and making a social identity salient shifts people's patience and risk-taking (Benjamin et al., 2010). Bloch and Kranton (2025) show that when people cannot tell whom they are dealing with, minority members tend to adopt the majority's behavior, a "tyranny of the majority." A voicemail greeting is a natural case of that uncertainty, since it is recorded for callers the speaker cannot know in advance. Moving between both modes may be one response to that uncertainty.

Communication across groups has long been seen as a source of labor market frictions (Lang, 1986). Recent evidence suggests these frictions shape who ends up in social-skill-intensive jobs (Noray and Noray, 2026), and informal interaction with managers accounts for a large share of the gender gap in promotion (Cullen and Perez-Truglia, 2023). Communication skills, verbal and nonverbal, have always mattered at work, but economic research has mostly studied hard skills, which are easier to measure. Social skills have grown in value (Deming, 2017; Green, 2024), and as Roslansky and Raman (2026) argue, soft skills like communication may matter more still as AI takes on routine and technical work. If AI substitutes more readily for technical work than for interaction, the human edge shifts toward communication. I conjecture that this raises the stakes of conforming: as communication gains value, out-group workers may face more pressure to adopt the in-group's style, and behavior like mode-switching may become more common and more consequential. That makes it worth measuring.

In a male-dominated work environment, I find robust evidence of mode-switching among women, especially junior women. Why they do it, whether it is a choice, and what it costs or gains them are the next questions, and they can now be studied at scale. Until recently, firms that wanted to understand how people interact had little beyond surveys and focus groups. Speech is already treated as a measurable signal elsewhere: medical researchers use fine-grained features such as pitch and pauses to flag genetic risk for Alzheimer's (Belluck, 2026; Tavakoli et al., 2026). Workplace data remain confidential, but organizations need far fewer outside experts and smaller budgets to study themselves, though new measures still need checking in each setting: AI tools make it cheap to collect and analyze calls, meetings and other unstructured records of work, and to measure subtle behavior directly.[12] Used with care, such analysis can show how teams interact and how identity shapes those interactions, and help match workers to tasks and teams in ways that raise both productivity and well-being.

References

Aghion, Philippe. "We should enhance manual jobs and soft skills." Interview, *Nobel Prize Outreach*, 2025.

Akerlof, George A. "A theory of social custom, of which unemployment may be one consequence." *Quarterly Journal of Economics* 94(4), (1980): 749–775.

Akerlof, George A., and Rachel E. Kranton. "Economics and identity." *The Quarterly Journal of Economics* 115, no. 3 (2000): 715–753.

Apple, William, Lynn A. Streeter, and Robert M. Krauss. "Effects of pitch and speech rate on personal attributions." *Journal of Personality and Social Psychology* 37, no. 5 (1979): 715–727.

Argyle, Michael. "Non-verbal communication in human social interaction." In *Non-verbal communication*, edited by Robert A. Hinde, 243–270. London: Cambridge University Press, 1972.

Auer, Peter, ed. *Code-switching in conversation: Language, interaction and identity*. London: Routledge, 1998.

[12] Building the dataset for this paper took years of writing code to link phone directories, call-management software and pitch tracking, and then of cleaning and analyzing the recordings. Today's AI tools can streamline much of that work and cut its cost.

Austen-Smith, David, and Roland G. Fryer Jr. "An economic analysis of 'acting white'." *Quarterly Journal of Economics* 120, no. 2 (2005): 551-583.

Baik, Bok, Alex G. Kim, David S. Kim, and Sangwon Yoon. "Vocal delivery quality in earnings conference calls." *Journal of Accounting and Economics* 80(1), (2025): 101763.

Banse, Rainer, and Klaus R. Scherer. "Acoustic profiles in vocal emotion expression." *Journal of Personality and Social Psychology* 70, no. 3 (1996): 614–636.

Baus, Cristina, Phil McAleer, Katherine Marcoux, Pascal Belin, and Albert Costa. "Forming social impressions from voices in native and foreign languages." *Scientific Reports* 9, no. 1 (2019): 1–14.

Belluck, Pam. "Subtle changes in speech may help predict dementia risk." *New York Times*, November 11, 2026.

Benjamin, Daniel J., James J. Choi, and A. Joshua Strickland. "Social identity and preferences." *American Economic Review* 100(4), (2010): 1913–1928.

Biddle, Jeff E., and Daniel S. Hamermesh. "Beauty, productivity, and discrimination: Lawyers' looks and lucre." *Journal of Labor Economics* 16(1), (1998): 172–201.

Biskupic, Joan, Janet Roberts, and John Shiffman. "Echo chamber: A small group of lawyers and its outsized influence at the US Supreme Court." *Reuters* (2014). *https://www.reuters.com/investigates/special-report/scotus*.

Bloch, Francis, and Rachel Kranton. "True colors: Authenticity in social and economic interactions." Working paper, Duke University, December 2025.

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.

Burgoon, Judee K., David B. Buller, and William G. Woodall. *Nonverbal Communication: The Unspoken Dialogue*. 2nd ed. New York: McGraw-Hill, 1996.

Bursztyn, Leonardo, Thomas Fujiwara, and Amanda Pallais. "'Acting Wife': Marriage Market Incentives and Labor Market Investments." *American Economic Review* 107, no. 11 (2017): 3288-3319.

Chen, Daniel, Yosh Halberstam, and Alan C.L. Yu. "Perceived masculinity predicts US Supreme Court outcomes." *PLoS One* 11, no. 10 (2016): e0164324.

Chozick, Amy. "Liz Holmes wants you to forget about Elizabeth." *New York Times*, May 7, 2023.

Creel, Sarah C., and Micah R. Bregman. "How talker identity relates to language processing." *Language and Linguistics Compass* 5, no. 5 (2011): 190–204.

Cullen, Zoë B., and Ricardo Perez-Truglia. "The old boys' club: Schmoozing and the gender gap." *American Economic Review* 113(7), (2023): 1703–1740.

Das, Jayoti, and Stephen B. DeLoach. "Mirror, mirror on the wall: The effect of time spent grooming on earnings." *Journal of Socio-Economics 40(1)*, (2011): 26–34.

Deb, Partha, and Pravin K. Trivedi. "Demand for medical care by the elderly: A finite mixture approach." *Journal of Applied Econometrics* 12, no. 3 (1997): 313–336.

Deb, Partha. "fmm: Stata Module to Estimate Finite Mixture Models." Boston College Department of Economics, Statistical Software Components s456895 (2012).

Deming, David J. "The growing importance of social skills in the labor market." *Quarterly Journal of Economics* 132(4), (2017): 1593–1640.

Ekman, Paul, Wallach V. Friesen, and Klaus R. Scherer. "Body movement and voice pitch in deceptive interaction." *Semiotica* 16, no. 1 (1976): 23–27.

Ewertz, Jonas, Charlotte Knickrehm, Martin Nienhaus, and Doron Reichmann. "Listen closely: Measuring vocal tone in corporate disclosures." *Journal of Accounting Research* 64(1), (2026): 229–277.

Fisher, Rhys. "Krisp launches customer accent conversion for global contact centers." *CX Today*, March 11, 2026.

Formisano, Elia, Federico De Martino, Milene Bonte, and Rainer Goebel. "'Who' is saying 'what'? Brain-based decoding of human voice and speech." *Science* 322, no. 5903 (2008): 970–973.

Fryer Jr, Roland G., and Paul Torelli. "An empirical analysis of 'acting white'." *Journal of Public Economics* 94, no. 5-6 (2010): 380-396.

Gardner-Chloros, Penelope. *Code-switching*. London: Cambridge University Press, 2009.

Giles, Howard, and Peter Powesland. "Accommodation theory." In *Sociolinguistics: A reader*, edited by Nikolas Coupland and Adam Jaworski, 232–239. Palgrave, London, 1997.

Green, Andrew. "Artificial intelligence and the changing demand for skills in Canada: The increasing importance of social skills." OECD Artificial Intelligence Papers No. 17. Paris: OECD Publishing, 2024.

Grogger, Jeffrey. "Speech patterns and racial wage inequality." *Journal of Human Resources* 46, no. 1 (2011): 1–25.

Grogger, Jeffrey. "Speech and wages." *Journal of Human Resources* 54(4), (2019): 926–952.

Heller, Monica, ed. *Codeswitching: Anthropological and sociolinguistic perspectives*. Vol. 48. Berlin: Walter de Gruyter, 1988.

Heller, Monica. "The politics of codeswitching and language choice." *Journal of Multilingual & Multicultural Development* 13, no. 1-2 (1992): 123–142.

Imhof, Margarete. "Listening to voices and judging people." *The International Journal of Listening* 24, no. 1 (2010): 19–33.

Jeffries, Michael P., Travis L. Gosa, and Erik Nielson. "The king's english: Obama, Jay Z, and the science of code switching." In *The Hip Hop & Obama Reader*, edited by Travis L. Gosa and Erik Nielson, 243–261. New York: Oxford University Press, 2015.

Klofstad, Casey A., Rindy C. Anderson, and Susan Peters. "Sounds like a winner: Voice pitch influences perception of leadership capacity in both men and women." *Proceedings of the Royal Society B: Biological Sciences* 279, no. 1738 (2012): 2698–2704.

Lang, Kevin. "A language theory of discrimination." *Quarterly Journal of Economics* 101(2), (1986): 363–382.

Lee, Soohyung, and Keunkwan Ryu. "Plastic surgery: Investment in human capital or consumption?" *Journal of Human Capital* 6(3), (2012): 224–250.

Leongómez, Juan David, Viktoria R. Mileva, Anthony C. Little, and S. Craig Roberts. "Perceived differences in social status between speaker and listener affect the speaker's vocal characteristics." *PLoS One* 12, no. 6 (2017): e0179407.

Mathias, Samuel R., and Katharina von Kriegstein. "How do we recognise who is speaking?" *Frontiers in Bioscience* 6 (2014): 92–109.

Mayew, William J., Christopher A. Parsons, and Mohan Venkatachalam. "Voice pitch and the labor market success of male chief executive officers." *Evolution and Human Behavior* 34, no. 4 (2013): 243–248.

McCluney, Courtney L., Kathrina Robotham, Serenity Lee, Richard Smith, and Myles Durkee. "The costs of code-switching." *Harvard Business Review* (2019).

McGettigan, Carolyn, and Sophie K. Scott. "Cortical asymmetries in speech perception: What's wrong, what's right and what's left?" *Trends in Cognitive Sciences* 16, no. 5 (2012): 269–276.

Myers-Scotton, Carol. *Social motivations for codeswitching: Evidence from Africa*. New York: Oxford University Press, 1995.

Noray, Kadeem, and Savannah Noray. "Workplace communication as systemic disadvantage." Harvard Business School Working Paper 26-059, 2026.

O'Hair, Dan, and Michael J. Cody. "Machiavellian beliefs and social influence." *Western Journal of Communication* 51, no. 3 (1987): 279–303.

Pesendorfer, Wolfgang. "Design innovation and fashion cycles." *American Economic Review* 85(4), (1995): 771–792.

Peters, Alex. "Unspoken phenomenon: why women are deepening their voices in the workplace." *Dazed Digital*, March 10, 2022.

Pisanski, Katarzyna, Emanuel C. Mora, Annette Pisanski, David Reby, Piotr Sorokowski, Tomasz Frackowiak, and David R. Feinberg. "Volitional exaggeration of body size through fundamental and formant frequency modulation in humans." *Scientific Reports* 6, no. 1 (2016): 1–8.

Rask, Mathias, and Frederik Hjorth. "Partisan conflict in nonverbal communication." *Political Science Research and Methods* 14(3), (2026): 771–788.

Rittmann, Oliver. "Legislators' emotional engagement with women's issues: Gendered patterns of vocal pitch in the German Bundestag." *British Journal of Political Science* 54(3), (2024): 937–945.

Roslansky, Ryan, and Aneesh Raman. *Open to Work: How to Get Ahead in the Age of AI*. New York: Harper Business, 2026.

Royle, Orianna Rosa. "Bad News, Graduates: LinkedIn Says the Most In-Demand Skills of 2024 Can't Be Learned in a Textbook." *Fortune*, February 8, 2024. https://fortune.com/2024/02/08/linkedin-most-in-demand-skills-of-2024/

Scott, Sophie K. "From speech and talkers to the social world: The neural processing of human spoken language." *Science* 366, no. 6461 (2019): 58–62.

Silverman, Bernard W. "Using kernel density estimates to investigate multimodality." *Journal of the Royal Statistical Society: Series B (Methodological)* 43, no. 1 (1981): 97–99.

Smith, David RR, and Roy D. Patterson. "The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age." *The Journal of the Acoustical Society of America* 118, no. 5 (2005): 3177–3186.

Starecheski, Laura. "Can changing how you sound help you find your voice?" *All Things Considered*, NPR, October 14, 2014.

Tavakoli, Maryam, Mehrdad Dadgostar, Jordan R. Green, David H. Salat, Steven E. Arnold, Kathryn P. Connaghan, Brian Richburg, Nelson V. Barnett, Mariam Tkeshelashvili, and Marziye Eshghi. "Remote identification of APOE-ε4 carrier status in cognitively normal adults via speech acoustics: Cross-sectional observational study." *JMIR Formative Research* 10 (2026): e89830.

Tigue, Cara C., Diana J. Borak, Jillian J.M. O'Connor, Charles Schandl, and David R. Feinberg. "Voice pitch influences voting behavior." *Evolution and Human Behavior* 33, no. 3 (2012): 210–216.

Weston, Philip S.J., Michael D. Hunter, Dilraj S. Sokhi, Iain D. Wilkinson, and Peter W.R. Woodruff. "Discrimination of voice gender in the human auditory cortex." *NeuroImage* 105 (2015): 208–214.

Wiggers, Kyle. "Sanas taps AI to change call center workers' accents in real time." *TechCrunch*, February 19, 2025.

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.