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

Yosh Halberstam

March 2023

#### Abstract

This study breaks ground by unlocking a new type of microbehavior, which connects to a phenomenon most often associated with out-groups and popularly referred to as codeswitching. To date, studies of human voice comprise a limited number of subjects, preventing researchers from detecting robust distributional patterns in human speech. Using thousands of voicemail greetings of lawyers at top U.S. law firms–a male dominated work environment–I show that 36 percent of females alternate between a primary frequency of about 200 Hz and a secondary frequency of about 100 Hz. The latter mode accounts for 8.5 percent of the signal and is coextensive with the unique male voice frequency mode. Likewise, the tendency to *mode-switch* is stronger among junior than among senior females, and survey data suggest that human listeners can detect the bimodality and perceive this group of females to be lower ranking. Evidence from auxiliary data sources provides external validity for the phenomenon. These findings suggest that the benefits out-group members typically accrue from conforming to in-group norms are unlikely to be realized for workers who oscillate.

<sup>\*</sup> 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, 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.

# 1. Introduction

Conforming to norms in a heterogeneous workplace has been shown to have unequal consequences for different types of workers: workers whose natural behavior is similar to that prescribed by norms are less affected than those with dissimilar behavior. In recent years, however, there has been growing concern about the built-in disadvantages out-groups (i.e., minorities, or marginalized groups) face at work. Specifically, because norms in the workplace are often driven by the in-group, they are relatively more costly for the out-group to follow and further hamper these workers from reaching their labor market potential. This work seeks to highlight these disparities and the need to focus on policies that level the playing field in a diverse workplace.

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 human 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). Although research on conformity to social norms at work is not new, an increasing number of anecdotes, especially among African American and female workers in the U.S., describe a new type of behavioral response to these pressures that neither uniquely conforms to in-group nor out-group norms.

In this first large-scale study of the human voice, recordings of workers in a maledominated work environment show that about one third of females subtly alternate between two distinct frequencies—100 Hz and 200 Hz—within a fraction of a second. Combined with auxiliary data I collected from other work contexts, there is some evidence showing a higher prevalence of

*mode-switching* among females in more vulnerable positions. Likewise, survey results confirm that human listeners perceive females with unimodal speech patterns to be more dominant and high-ranking than females with bimodal speech patterns. In contrast, I find no evidence of bimodal speech patterns among male workers. Although adapting one's behavior can foster better communication among diverse individuals, this unilateral practice of assimilation appears to impose an unfair burden on female workers.

These findings connect to a phenomenon, popularly referred to as *code-switching*, which has gained traction, reviving a literature at the intersection of linguists and anthropology.<sup>1</sup> Unlike conforming to the norms of a single group, workers signalling deference to multiple groups endure additional psychological costs of keeping their identity hanging in the balance (McCluney et al., 2019). Put differently, identity-switching at work may result from out-group employees experiencing competing pressures. This non-standard type of conformism, whereby workers briefly yet regularly express multiple social identities in a single utterance is significantly more challenging to detect, let alone empirically document. This paper begins to fill this important gap.

Economists have studied the role of assimilation in identity formation. Specifically, Austen-Smith and Fryer (2005) developed a two-audience signalling model to explain the pressures faced and choices made by out-group members when in-group norms dictate behavior that conflicts with out-group norms. In their paper, one may only conform to the norms of a single group. In contrast, mode-switching can be seen as a hybrid: marginalized workers shifting between their native out-group and the in-group market norms, thereby simultaneously, but not fully, conforming to divergent norms. Several recent papers have examined the role of social

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

influence in under-achievement of out-groups in educational attainment (Fryer and Torelli, 2010) and professional identity choices (Bursztyn et al., 2017). Benjamin et al. (2010) show in a lab experiment that making ethnic identity salient causes risk and time preferences of subjects to conform to common stereotypes. Evidence from these papers implies that conflicting social influences puts members of the out-group at a relative disadvantage. However, even the benefits that typically come with group membership are unlikely to be realized for workers who oscillate.

In a different strand of the economics literature, the roots of discrimination are hypothesized to originate in communication difficulties across groups (Lang, 1986). To the extent that vocal similarity is important for maintaining social relationships, it is plausible that individuals may face pressure to conform (and be penalized for not conforming) to dominant market norms concerning voice. Other research on market norms posits that they may reflect short-lived fads and fashion cycles (Pesendorfer, 1995) or, alternatively, persistent social customs (Akerlof, 1980). Because such dynamics are governed by in-group preferences that are not aligned with those of out-group members, they may generate uncertainty and thus exacerbate the challenges borne by out-group members.

Together, these various literatures provide context and motivation for investigating voice frequency as a possible arena for codeswitching. Voice has been established as a non-trivial factor in how human beings size up one another. And, unlike alternative examples of nonverbal behavior, such as how individuals dress or present themselves at work, voice frequency is a quantifiable measure of non-verbal behavior that is unrelated to key confounding factors, such as income.

### 2. Background and Preliminary Findings

The evidence on neural processing of vocal cues has improved markedly in recent years (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 new research now provides a rigorous framework for a large literature in the social sciences on snap judgements (Creel and Bregman, 2011; Weston et al., 2015), which draws connections between vocal cues and listeners' subjective perceptions of a speaker's attributes (Imhof, 2010; Baus et al., 2019; Grogger, 2011; Chen et al., 2016; Buller et al., 1996; O'Hair and Cody, 1987). In contrast, measurement of human voice production has remained, for the most part, static.

To date, studies of human voice comprise a limited number of subjects (e.g., Leongómez et al. 2017; Pisanski et al., 2016; Smith and Patterson, 2005; Banse and Scherer, 1996), preventing researchers from detecting robust distributional patterns in human speech. Specifically, human anatomy enables one to manipulate the vocal cords finely and rapidly; however, existing studies have largely focused on a person's *mean* voice frequency (Klofstad et al., 2012; Apple et al., 1979; Tigue et al., 2012; Ekman et al., 1976; Mayew et al., 2013), treating pitch as a unimodal characteristic of speech. This study shows that *modal* frequencies can illustrate 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. This suggests that male vocal frequencies have a dominant influence in the workplace.

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 account for about 25 percent of total revenues in the legal services industry.<sup>2</sup> Firm-level descriptive statistics gathered from external sources of data about law firms are presented in Table 1A.<sup>3</sup> 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 homogenous 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.

The data assembly entailed scraping the phone directories from each firm's webpage, using a call management software to call each phone, and recording the voicemail greeting once the call was connected. The calls were made in early 2018, primarily during weekend nights to maximize the chances of reaching the lawyers' voicemails. Each 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

<sup>&</sup>lt;sup>2</sup> 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).

<sup>&</sup>lt;sup>3</sup> 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.

Compass database, a leading directory of lawyers, into the dataset by phone number and lawyer name. Table 1B 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 1A 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 the Online Appendix 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 ( $\mu$ ) and length (*L*):

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

The basic approach involves finding a set of frequency candidates that produce the highest autocorrelation, adjusting for the fact that the autocorrelation function mechanically favors lower frequencies (e.g., a signal that repeats itself every 5 milliseconds, also repeats itself every 10, 15, 20, and so on milliseconds).

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, given that not all workers personally record their voicemail greeting, I combine machine learning techniques with biographical information about the lawyers as well as verbal and nonverbal characteristics of the greeting to isolate greetings that were self-recorded 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 1C 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 1A shows this subsample of frequency estimates for female and male lawyers separately. The horizontal axis in the figure represents time lapsed 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. This is the key finding that motivates the analysis in the study.

### 3. Between- or Within-Person Bimodality?

Does this secondary mode reflect females that vocalize like males or bimodal vocalization of females (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, ..., 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:

$$\int_{0}^{F_0^{i,p}} h_i(\nu) \mathrm{d}\nu \le p/100 \tag{2}$$

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 1B 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 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. 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 as conservative since the bootstrapped samples are drawn from the critical density only and tends to underestimate the true number of modes.

Table 2A 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*-value = 0).<sup>4</sup>

<sup>&</sup>lt;sup>4</sup> The results cannot reject that the number of modes is 2 or more; 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.

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

#### 4. Is Mode-Switching Widespread?

My empirical approach to answering this question 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:

$$h_{i}() = \sum_{k=1}^{g} \pi_{i,k} N(\mu_{i,k}, \sigma_{i,k}^{2})$$
(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 2B 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.

In Figure 1C, I present clip-level estimation results of the lowest frequency mode location (i.e.,  $\min_{k} \hat{\mu}_{i,k}$ ). 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 solid line along with the predicted individual normal distributions in the mixture. The implied cut-off of 115 Hz between both groups indicates that 36 percent (0.7 delta method standard error) of female lawyers (Group 1) have low mode estimates more than 80 Hz below the *mean* frequency estimate.

For robustness, Table 2C shows how the estimated share of Group 1 (bimodal) and Group 2 (unimodal) clips vary with alternative specifications for estimating frequency modes. Rows 1-6 use a 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 2B in showing that g = 5 maximizes the model fit.

Using the demeaned percentiles described above, I plot the histograms of each group of female lawyers in Figure 1D. The histogram of female lawyers associated with the high values of low mode estimates (Group 2) shows no sign of bimodality, whereas the histogram of Group 1 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 (0.2 delta method standard error, adjusted for clip-level clustering) of the 5-component mixture density.

### 5. Is Mode-Switching Context-Specific?

So far, this article has documented a new type of expression among female lawyers. I next examine whether the prevalence of this pattern varies by worker or firm characteristics. Subsequently, I analyze data from several additional sources to explore the external validity of my findings in other workplace contexts. Using the same mixture model methodology described earlier, I summarize the estimation results in Figure 2. In this figure, 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 indicated in parenthesis.

In Figure 2A, 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 differentially interact with clients and judges relative to non-litigators; however, I find no difference between both 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. Clearly, there are many reasons that can drive this difference, but years since graduation from law school, a common proxy for both experience and age, is not among them. The correlation between the individual low mode location estimates and the residuals from running these estimates on graduation-year, firm, title, and litigator fixed effects is 0.985.

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 homogenous sector. For example, based on 2016 headcounts, the average share of females in these firms was 0.36 with a standard deviation of 0.03 indicating limited between-firm variation (Table 1A). Likewise, also with a standard deviation of 0.03, the average share of female 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 2B relative to Figure 2A. 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 subsequently promoted the following year. 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 may instead indicate a behavioral response to a change in the workplace environment following promotion.

One context that forces a worker to change their voicemail greeting is switching jobs. One year 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, slightly more than one third of voicemail greetings from each period of collection are bimodal (Group 1).<sup>5</sup>

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 sentences in an equally male dominant 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 (41 percent) of the middle (end) argument sample is estimated to have a secondary mode; the differences are statistically insignificant. Overall, the findings

<sup>&</sup>lt;sup>5</sup> More details about this sample are available in the Online Appendix.

identify mode-switching as a phenomenon outside the office and beyond the introductory sentence.

The next set of results comes from two female dominant 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, my findings on bimodal vocalization of female workers externalize beyond lawyers and the first few seconds of speech, suggesting the existence of a widespread phenomenon among females in the labor market.

# 6. Perceptions from Bimodal Vocalization

The findings above indicate that a significant proportion of female workers mode-switch. 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 approval rating of 99 percent or more. The workers ranged from age 18 through 74 with the median age being 36.

I paired 250 clips of female lawyers from Group 1 with 250 clips of female lawyers from Group 2, 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. 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 2C. 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 red line means that workers perceived the attribute to resonate more strongly with lawyers from Group 2 than Group 1. 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. Females

perceive unimodal vocalization approximately one quarter of a standard deviation more dominant and senior than bimodal vocalization. For male listeners, the perceived difference is about half that size.

In sum, results from this survey suggest that humans use the acoustic signal to inform their perceptions of the speaker. Group 2 lawyers are perceived as significantly more dominant and senior than Group 1 lawyers.<sup>6</sup>

# 7. Mean Voice Frequency and Deep Voice Premium

To provide context for my findings, I provide a description of 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, Leongomez et al. (2017) show that

<sup>&</sup>lt;sup>6</sup> For robustness, I explore the role of verbal content but do not find meaningful differences in the choice of words by both groups. See the Online Appendix for this comparison.

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).

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 3A 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 centered around 200 Hz, the female vocal mode, with no evidence of females permanently adopting a male voice frequency. 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.<sup>7</sup> Although the mean voice frequency is strongly positively correlated with being an Associate (Figure 3C on the left) relative to a Partner, the relationship becomes insignificant when controlling for years of experience (Figure 3C on the right), which is also a close proxy for age.

## 8. Conclusion

Conforming to norms in the workplace is a greater task for out-group individuals because market norms are often driven by the preferences of the in-group (Akerlof and Kranton, 2000).

<sup>&</sup>lt;sup>7</sup> 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 et al., 2011), and an empirical challenge to exploring beauty-based codeswitching is the difficulty of measuring beauty in an objective manner across time and space.

For example, in a male-dominated workplace, females may experience greater modifications to their behavior and more pressures to "fit in" than males. The manifestation of these pressures in behaviors can be subtle and challenging to detect.

Just as language can reflect identity (Auer, 1998; Myers-Scotton, 1995), so can nonverbal vocalization (Argyle, 1972). This new evidence on voice frequency mode-switching by female workers connects to the social phenomenon of codeswitching, a concept that originated in the linguistics literature (Gardner-Chloros, 2009; Heller, 1992). More recently, codeswitching has been extended to describe a subtle and brief form of out-group expression (e.g., Jeffries et al., 2015), possibly to signal the recognition of, deference to, or resonance with in-group norms. Unlike other accommodative behaviors (Giles and Powesland, 1997), codeswitching preserves the integrity of each underlying norm and the prescribed conventions associated with it (Heller, 1988).

In the male dominated work environment, I find robust evidence for mode-switching behavior among female workers, particularly those in relatively vulnerable positions. A growing number of anecdotes suggest that this unilateral form of accommodative behavior comes at a cost with little evidence of a benefit. More than celebrating diversity, awareness of these inequities and the unique cultural capital one brings to the workplace are necessary.

# References

- Akerlof, Geroge 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.
- Austen-Smith, David, and Roland G. Fryer Jr. "An economic analysis of 'acting white'." *Quarterly Journal of Economics* 120, no. 2 (2005): 551-583.
- 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.
- 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.*
- Boersma, Paul. "Accurate short-term analysis of the fundamental frequency and the harmonicsto-noise ratio of a sampled sound." *Proceedings of The Institute of Phonetic Sciences* 17, no. 1193 (1993): 97–110.
- Boersma, Paul. "Praat, a system for doing phonetics by computer." *Glot International* 5, no. 9 (2001): 341–345.

- Burgoon, Judee K., David B. Buller, and William G. Woodall. *Nonverbal Communication: The Unspoken Dialogue*. 2<sup>nd</sup> 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.
- Creel, Sarah C., and Micah R. Bregman. "How talker identity relates to language processing." *Language and Linguistics Compass* 5, no. 5 (2011): 190–204.
- Das, Jayoti, and Stephen B. De Loach. "Mirror, mirror on the wall: The effect of time spent grooming on earnings." *Journal of Behavioral and Experimental 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).
- Ekman, Paul, Wallach V. Friesen, and Klaus R. Scherer. "Body movement and voice pitch in deceptive interaction." *Semiotica* 16, no. 1 (1976): 23–27.
- 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.
- Grogger, Jeffrey. "Speech patterns and racial wage inequality." *Journal of Human Resources* 46, no. 1 (2011): 1–25.
- Heller, Monica, ed. *Codeswitching: Anthropological and sociolinguistic perspectives*. Vol. 48. Berlin: Walter de Gruyter, 2010.
- 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.
- Klatt, Dennis H. "Discrimination of fundamental frequency contours in synthetic speech: Implications for models of pitch perception." *The Journal of the Acoustical Society of America* 53, no. 1 (1973): 8–16.
- 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.
- Kollmeier, Birger, Thomas Brand, and Bernd Meyer. "Perception of speech and sound." In *Springer handbook of speech processing*, edited by Jacob Benesty, M. Mohan Sondhi, and Yiteng Arden Huang, 61–82. Berlin: Springer, 2008.
- 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.
- Lehiste, Ilse. Suprasegmentals. Cambridge: MIT Press, 1970.
- 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.
- 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*, (1995): 771–792.
- 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.
- 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.
- 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.
- Vogel, Adam P., Paul Maruff, Peter J. Snyder, and James C. Mundt. "Standardization of pitchrange settings in voice acoustic analysis." *Behavior Research Methods* 41, no. 2 (2009): 318-324.
- 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.



**Figure 1. Bimodal voice frequency of female lawyers. A.** 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. B. 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. **C.** The location of a low mode is individually estimated for each of 6,399 verified self-recorded greetings. **D.** Based on this group classification, histograms of the demeaned frequency estimates are shown for each group of female lawyers. The results show a unimodal density for Group 2 and a bimodal density for Group 1 with a low secondary mode accounting for 8.5% (standard error 0.20) of the density.



ю

1 0.75 0.5 0.25 0 0.25 0.5 0.75 1

Standard Deviations

**Figure 2.** The incidence and impressions of females with bimodal voice frequency. A. This figure displays the estimated share (with associated 95% confidence intervals) of recordings in each sample that contains 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 (Main), the number of clips in each subsample is indicated in parentheses. Six unranked firms were omitted from the Women in Law comparison. See text for description of subsamples. **B.** This figure uses supplemental data (Auxiliary). The number of clips in each subsample is indicated in parentheses. See text and the Online Appendix for more details about these samples. **C.** 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 red vertical line is the perceived difference between Group 1 and Group 2 in terms of standard deviations of each attribute ratings. Survey results suggest that humans, especially females, perceive unimodal vocalization significantly more dominant and high ranking than bimodal vocalization.

0.75 0.5 0.25 0 0.25 0.5 0.75 1

Standard Deviations



**Figure 3.** Mean voice frequency, experience, and seniority among female lawyers. **A.** 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 normal distribution around approximately 200 Hz, the primary female vocal mode. **B.** 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. **C.** 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 plot on the left shows a strong positive correlation between the mean frequency and the likelihood of being junior; however, as seen on the right, the relationship becomes significantly weaker when controlling for experience.

A: Firm Descriptive Statistics							
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		

B: Lawyers by Title and Gender								
	Asso	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
C: Likelihood of Self-Recorded Voicemail Greeting								
	Prob > 0.95 Prob > 0.50		> 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

**Table 1. Descriptive statistics. A.** 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. **B.** This table presents the number of recordings in the main dataset of lawyer voicemail greetings by lawyer gender and job title. **C.** 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 Online Appendix for details on the machine learning classification for the probability that a voicemail greeting was self-recorded by the lawyer.

Ho	Critical Bandwidth	Change	Silverman	Mode Location			
0	(HZ)	(%)	<i>p</i> -value	(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			

A: Silverman Test for Number of Modes

B: FMM Optimal Number of Components							
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

B: FMM	Ontimal	Number	of (	Components
D, I W W	Optimar	Trannoer		

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

Table 2. Robustness and measurement of modes. A. 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 selfrecorded 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). B. This table shows aggregate model fit for a variety of mixture models using 100 demeaned percentiles of a random sample of 1,000 from all 6,399 clips in the sample of verified self-recorded female lawyers. The results show that q = 5 maximizes model fit according to Bayesian information criterion (BIC), and the marginal improvement from q = 5 onwards is below 1% according to Akaike's information criterion (AIC). C. This table shows how the estimated share of Group 1 (bimodal) and Group 2 (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 1,000 female lawyers described above. 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.