University of Toronto
Department of Economics

Working Paper 636

Voice at Work

By Yosh Halberstam

April 08, 2019

Voice at Work*

Yosh Halberstam†

March 20, 2019

Abstract

In the first large-scale study on voice, audio data on lawyers at the top U.S. law firms–a male dominated work environment–show that female lawyers alternate between two voice frequency modes: a primary female mode at about 200 Hz as well as a secondary female mode at about 100 Hz that is coextensive with the primary (and only) male voice frequency mode. This tendency is stronger among female associates than among female partners, and does not replicate for male lawyers or female assistants. Evidence of differences driven by firm heterogeneity is comparatively insignificant, indicating market-wide trends in workplace behavior.

Keywords: labor markets; gender identity; social norms; codeswitching; voice frequency
JEL Classification: D91, J16, J44, M14; Z10

*I thank Heski Bar-Issac, Jonathan Beauchamp, Dan Benjamin, Dwayne Benjamin, Tim Besley, Alberto Bisin, Arthur Blouin, Gustavo Bobonis, Martin Burda, Daniel Chen, Yoko Cohen, Chris Cotton, Peter Cziraki, Stefano DellaVigna, Shari Eli, Hanming Fang, Raquel Fernandez, May Friedman, Sarah Fulford, Paola Giuliano, Jim Goldman, Marco Gonzales-Navarro, Christian Gourieroux, Joshua Halberstam, Yoram Halevy, Jonathan Hall, Brent Hickman, Mitch Hoffman, Tanjim Hossain, Emir Kamenica, Sarah Kaplan, Brian Knight, Kory Kroft, Nico Lacetera, Kevin Lang, Steve Lehrer, Danielle Li, Bentley MacLeod, Rob McMillan, Peter Morrow, Sendhil Mullainathan, Phil Oreopoulos, Daniele Paserman, Imran Rasul, Heather Sarsons, Emily Satterthwaite, Michel Serafinelli, Aloysius Siow, Jeffrey Smith, Colin Stewart, Eduardo Souza Rodrigues, Hugo Troiano, Hans-Joachim Voth, and Alan Yu, as well as seminar participants at Bar-Ilan University, Colgate University, Queen's University, York University, University of Rochester, and the University of Toronto. Raheeb Dastagir, Katherine Lavrova, David Rhodes, Sida Shuai, Fangyu Yan and Jinguang Yang provided outstanding research assistance. This work was partially supported by the Social Sciences and Humanities Research Council of Canada.

†Department of Economics, University of Toronto, yosh.halberstam@utoronto.ca. Please check my website for the latest draft.

Monica Hanna, a tough litigator in New York City, is about 5 feet tall and has a high voice... one of the partners at her firm assessed a presentation she gave by telling her: "Your voice is very high."

"And then he didn't say anything else," says Hanna. "He didn't have any other comment to make about my presentation at all." So, when her high voice came up again in an evaluation about a year ago, she decided to try to change this thing about herself that most of us think of as unchangeable.

"I want to be taken more seriously," she says, "from the first words out of my mouth to the last."

–Interview by Laura Starecheski[1]

1 Introduction

This paper represents the first large-scale study of voice in the workplace. An extensive literature in economics has examined the connection between labor market outcomes and worker characteristics. Within that, studies of discrimination have focused traditionally 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. This can occur when the 'out-group' (minority, or marginalized group) faces pressures to conform with a market norm that has been determined by the preferences of the 'in-group' (majority) (Akerlof and Kranton, 2000). Such norms can influence identity choices and behavior in the workplace. Under such conditions, observed correlations between outcomes and a malleable characteristic may result from the efforts of out-group workers to "fit in." Observable changes relating to the characteristic may thus convey useful information on the heterogeneous pressures faced by workers. For example, in a male-dominated workplace, females may experience more pressures than males to "fit in," resulting in greater modifications to their behavior. In this paper, I provide evidence consistent with this hypothesis.

Codeswitching is a term that originated in the linguistics literature, and traditionally referred to the practice of alternating between two or more languages in a single utterance. A hallmark of codeswitching is that it preserves the syntax and phonology of each underlying language. There is no "compromise" or "hybrid" language–there is simply switching between the two. More recently, the concept of codeswitching has been extended in the popular press to describe a behavior that members of the out-group use to signal their recognition of and deference to in-group norms, and possibly to cope with and mitigate the structural disadvantages they perceive. However, unlike

[1] Monica Hanna interviewed by Laura Starecheski in "Can Changing How You Sound Help You Find Your Voice?", *All Things Considered,* NPR, October 14, 2014. URL: https://www.npr.org/player/embed/354858420/356177267. Last accessed November 21, 2018.

other accommodative behaviors or forms of expression, codeswitching preserves the integrity of each underlying norm and the prescribed conventions associated with it.

To assess the evidence for codeswitching, I focus on the voice frequency of workers in a high-skilled male-dominated sector. Because codeswitching is bound up with a worker's identity choices, I exploit a ubiquitous feature of the professional workplace that requires workers to identify themselves using their voice: a voicemail greeting. Voice frequency is a useful characteristic for studying codeswitching behavior in this context because of two physiological properties. First, vocal cord tension is finely manipulable by the laryngeal muscles, and a higher or lower voice "pitch"–the perceived quality of the voice frequency–is the result of the exertion or relaxation of the vocal cord muscles. Second, because vocal cord length is proportional to body size, the average female and male voice frequencies diverge significantly from one another. I hypothesize that, in the work environment that I study, females (i.e., the out-group) may codeswitch between two divergent voice frequency modes in their voicemail greeting: one that is primary for females and the other that is primary for males.

Using data on 40,000 lawyers at the 100 most prestigious law firms in the U.S., this paper provides evidence broadly consistent with my hypothesis. The Vault 100 firms that I study account for about 25 percent of the total revenues of the legal services industry, and two-thirds of the lawyers in them are male.[2] At the partner level, three-quarters are male. 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 explore codeswitching in this context, I focus on within-greeting variation in voice frequency to assess whether there are distinct male and female voice frequency modes and, if so, whether there is evidence that some lawyers switch between the two modes. The codeswitching concept suggests that in-group workers–or, more broadly, workers not facing pressures to conform–are unlikely to engage in switching; that is, their voice frequency density functions will be largely unimodal. To study voice frequencies in an in-group context, and more specifically to assess whether there are different primary voice frequency modes for each gender group, I look at two groups of voicemail greetings: those that were self-recorded by a male lawyer and those that were recorded in the third person by an assistant on behalf of a lawyer. (Such assistants' voices were invariably female; accordingly, females can be construed as the in-group among assistants.) I find that both groups' voice frequency densities are strongly unimodal: for male lawyers, the mode is centered around 100 Hz. For female assistants, the mode is centered around 200 Hz–a very distinct difference. Finally, to assess codeswitching, I study the self-recorded voicemail greetings

[2]The Vault 100 is a ranking generated from survey responses of approximately 20,000 associate lawyers each year and is highly correlated with firm revenues. Based on the most recent Census NAICS (North American Industry Classification System) Code 5411 ("legal services"), total revenue in this industry is estimated at over 1/4 trillion dollars generated by approximately 200 thousand law offices across the US.

of the obvious out-group: female lawyers. Here, the data show that, in addition to the 200 Hz "female" primary voice frequency mode, female lawyers record a small share of their voicemail greeting using the "male" primary voice frequency mode at 100 Hz. Utilizing both the male and female frequency modes–which are almost 100 Hz apart from each other–within a few seconds of time requires fine modulation of the laryngeal muscles. Further, I find that this pattern is more distinctive among female associates than female partners, perhaps reflecting their relatively more precarious position within the out-group (i.e., more marginalized as compared to partners).

This evidence on voice frequency mode-switching behavior by female lawyers sheds new light on codeswitching in the workplace. To the extent that such behavior requires effort, the results imply that out-group workers are at a disadvantage. Growing evidence on this phenomenon from testimonies of marginalized workers across a wide range of professions indicate the toll codeswitching takes on worker wellbeing in the labor market overall.[3] This may further inhibit the performance and long term prospects of the out-group relative to the in-group.

Finally, to provide context for the results on codeswitching, I document cross-sectional patterns in the voice frequency of lawyers. Differences between mean voice frequencies (as distinguished from differences in the modes of voice frequency densities) reveal that while most female lawyers use the higher "female" voice frequency mode, a small group of female lawyers record their greeting using the lower "male" voice frequency mode. Moreover, the data suggest that female lawyers who record their greeting using primarily the "male" voice frequency mode enjoy higher rates of success as measured by the relatively higher share of partners in this group.

The paper proceeds as follows. In section 2, I discuss the relevant literature. Section 3 provides an overview of the vocal anatomy that determines one's voice frequency, and Section 4 describes methods used to collect and prepare the data for analysis. In Section 5, I discuss the concept of codeswitching and how it applies to the specific context of elite law firms. I describe my key findings in Section 6. In Section 7, I present cross-sectional voice patterns between lawyers, and address the question of whether unobserved lawyer characteristics can be predicted from voice data. Section 9 concludes.

2 Literature

The linguistics literature on codeswitching behavior initially focused on the use of more than one language in a single communication episode (Heller, 1988). More recently (as noted), the meaning

[3]For example, see https://www.girlboss.com/identity/code-switching-at-work. Codeswitching behavior has been gaining attention in the media (e.g., www.npr.org/sections/codeswitch/) as well as in the arts (e.g., Boots Riley's film "Sorry to Bother You"). Examples of codeswitching by political figures, such as Michelle and Barack Obama, demonstrate the skills required for post-black politicians, and recent books contemplate the possible costs this type of behavior entails (e.g., (Jeffries et al., 2015)).

of the term has been expanded to include a psychological phenomenon that transcends the study of linguistics (Vogt, 1954), and codeswitching now refers to subtle changes in the way people, typically members of marginalized groups, express themselves in different settings (Gardner-Chloros, 2009). Particularly in workplace settings in which market norms are non-native to members of marginalized groups, members of such groups may express themselves by codeswitching between market and native modes of behavior.

In economics, a large literature studies the connection between labor market outcomes and worker characteristics. However, unlike fixed worker attributes, such as race and sex, a malleable characteristic such as voice can be a reflection of labor market outcomes rather than a factor in determining them. Observed correlations between such a malleable characteristic and labor market outcomes may be the result of efforts on the part of workers to "fit in." Several recent papers have examined the role of social influence 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. These papers argue that conflicting social influences can disadvantage members of the out-group, who may face both scrutiny from their out-group peers and pressure to conform from the dominant in-group. The results from this paper can be seen as corroborating the presence of such pressures by documenting the out-group worker strategy of codeswitching.

Economists also have studied the role of assimilation in identity formation. Austen-Smith and Fryer (2005) develop a two-audience signalling model to explain the pressures faced by out-group members when in-group norms dictate behavior that conflicts with out-group norms. In their model, divergence between norms forces marginalized workers to conform to a single market norm. In contrast, codeswitching behavior can be seen as more of a hybrid: marginalized workers shift dynamically between their native out-group and the in-group market norms, thereby conforming at different times to different norms. Other research has proposed a role for culture cues (Cornell and Welch, 1996), social culture (Fang, 2001), and stereotypes in creating market discrimination (Coate and Loury, 1993). The results from this paper underscore the importance of such cultural hallmarks and connect with other research on conformity to standards of behavior at work (Bernheim, 1994), particularly with respect to the built "standard-setting" advantage enjoyed by members of the dominant in-group.

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). To the extent that 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.

Other work draws connections among communication, subordination, and discrimination (e.g., Lippi-Green (2012)). Thus, just as language can reflect political and social identity (Heller, 1992; Auer, 2013; Myers-Scotton, 1995), so can nonverbal cues, such as voice.[4] Because verbal cues are used for communicating about matters that are, in a sense, "external" to the speaker, identity may be more likely to be expressed using nonverbal cues (Argyle, 1972).

There is also a varied body of literature relating specifically to voice. Within economics, Grogger (2011) uses audio data from the 1997 National Longitudinal Survey of Youth (NLSY) to examine the connection between judgments about race–using voice–and earnings. He finds that African-American males misclassified by listeners as being white have significantly higher earnings than African-American males correctly identified by listeners as being black. In a similar vein, Mayew et al. (2013) uses data from quarterly conference call recordings of public companies listed in the S&P 1500 to find that CEOs with deeper voices manage larger companies.

In the literature on voice outside of economics, there is considerable evidence that emotion produces changes in respiration, phonation and articulation, which affect an individual's acoustic signal (see Banse and Scherer (1996)). A recent functional magnetic resonance imaging (fMRI) study found that brief exposure to voice frequency stimuli provides a strong cue for distinguishing between genders by a listener. Specifically, low voice frequencies consistently evoked a lower level of cortex activation in the listener, irrespective of the sex of the listener (Weston et al., 2015), suggesting that subtle switches between male and female modes of voice frequency–even in a very brief recorded interval–are detectable. In addition, voice frequency modulation has been observed cross-culturally, perhaps reflecting a universal "frequency code" (Ohala et al., 1984).[5] Other studies posit that perceptual biases based on the laws of physics, such as the intuition that larger objects resonate at lower frequencies, are likely to be universal precisely because they are determined by physics, not culture (Spence, 2011).[6] Smith and Patterson (2005) suggest that volitional voice frequency modulation may reflect an exploitation of perceptual biases linking low voice frequencies to large body size and dominance. Indeed, a number of surveys and lab experiments have found that listeners tend to judge speakers with deeper voices favorably. For example, speakers

[4] My findings also connect to communication accommodation theory (Giles and Powesland, 1997), where convergence or divergence in nonverbal communications is a strategic choice of the speaker.

[5] Several lab experiments document volitional voice frequency modulation by speakers. For example, in a simulated interview, Leongomez 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).

[6] For example, a recent lab experiment by Baus et al. (2019) showed that listeners were consistent in forming first impressions from voices independent of whether the spoken language was native or foreign to them.

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 competent, strong, trustworthy, truthful, and empathic, and perceived to possess greater leadership capacity (Klofstad et al., 2012; Apple et al., 1979; Buller et al., 1996). People also prefer to vote for candidates with low voices (Tigue et al., 2012) and associate higher-pitched voices with deception (Ekman et al., 1976; O'Hair and Cody, 1987).

Finally, there is a body of research that specifically addresses 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, and are more likely to work in the private sector. Along the same lines as voice, beauty also is mutable to some extent. 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.

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.

3 The Malleability of Voice Frequency

This section briefly explains what voice frequency is, and the human anatomy that renders it a malleable characteristic.

Voice frequency is determined by the vibration of the vocal cords. The vibrations of the vocal cords per unit of time determines the oscillation frequency of the soundwave produced by a human voice.

The vocal cords are located in the larynx (or "voice box"), which sits at the top of the trachea (see Figure A1 for graphical illustration of the anatomy). Because the anatomy of the larynx resembles the structure of other string instruments, Mersenne's laws (Mersenne, 1636) imply the following relationship between voice frequency and the anatomy of the larynx:

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

where T is the tension of the vocal cords, μ is the tissue mass, and L is the length of the vocal cords.[7] This relationship implies that, at any given moment in time, longer vocal cords, denser tissue mass, or less tension of the vocal cords result in a lower voice frequency (measured in units of hertz, or Hz).

Vocal cord length and tissue mass are relatively time-invariant, particularly in adulthood. In particular, vocal cord length is roughly proportional to body size. Thus, on average, the voice frequency of males is lower than that of females. Despite this variance of voice frequency by gender, one of the parameters in the voice frequency equation can be modulated–at least to some extent–at will: vocal cord tension. Humans can tighten or relax their laryngeal muscles to change the tension of their vocal cords. This results in changes to their voice's frequency. Less (more) tension results in a lower (higher) voice frequency, which is in turn perceived as having a lower (higher) "pitch." The ability to change the tension in one's vocal cords is the key feature that renders voice a malleable characteristic. Moreover, a substantial body of research has established that listeners can perceive very fine adjustments to voice frequencies. Early work by Lehiste (1970) places the Just Noticeable Difference (JND)–the smallest change in frequency that is perceptible by the average listener–at 1 Hz. Other estimates vary by context, but most studies place the JND of frequencies below 500 Hz at no more than 3Hz (Kollmeier et al., 2008; Klatt, 1973).

In sum, changes in voice frequencies are relatively easy to effect and easy to detect, but the starting points for those changes vary significantly by gender.

4 Data

4.1 Sample of Lawyers

My sample comprises 60,000 lawyers employed by 86 law firms that were listed at least once in the Vault 100 prestige rankings during the years 2016-2018.[8] These firms are, on average, male-dominated: two-thirds of all lawyers working at the firms (at both the associate and partner levels) are male, and one-third are female. Looking solely at the partnership level, the firms are even more heavily male: only one-fifth of partners are female. Firm-level descriptive statistics gathered from external sources of data about law firms (including but not limited to Vault.com; see notes to Table for details), are presented in Table 1. The average number of lawyers per firm is 973. The average profit per partner in these firms is \$1.53 million. On average, 36 percent of lawyers within a firm are female. Among partners, 21 percent is female. Among equity partners, 17 percent are female.

[7] Since the frequency is produced by the vibration whose nodes are the two ends of the vocal cords, if the length of the vocal cords is L, then the wavelength of the frequency is $2L$.

[8] I dropped firms that either had a live receptionist 24/7 (3 firms) or firms that had more than 90 percent of voicemail greetings recorded in third-person based on a pilot sample. See Appendix B for details on the pilot sample I used to determine the set of firms in my final sample.

The oldest firm in the dataset was established in 1792 and the youngest in 2014.

I collect voice frequency data from the voicemail greetings of the lawyers in my sample. This was feasible because, in the U.S. private legal services industry, it is standard for each lawyer in a firm to have his or her own phone number. When a call is not answered personally, a voicemail system picks up the call and the caller hears a recorded greeting.

To obtain phone numbers for the 60,000 lawyers in my sample, I used website scraping techniques. This resulted in 57,064 distinct lawyer phone numbers. Next, to record the voicemail greetings, I used Voice over Internet Protocol (VoIP) channels to call the phone numbers. Once a connection was initiated, an auto-dialer software program with call management features recorded the call. To maximize the chances of having the call sent to voicemail rather than being answered by a live lawyer or receptionist, the calls were made during off hours and weekends. Despite extensive efforts to ensure sufficient Internet bandwidth, approximately 5 percent of the lawyer voicemail recordings had to be dropped due to poor audio quality, typically as a result of accelerated voicemail greeting playback.[9]

4.2 Measuring the Voice Frequency of Lawyers

To assess codeswitching, it is necessary to estimate the voice frequency density functions–and observe the mode or modes thereof–for each individual lawyer. This requires lawyers' voice frequencies to be measured at different points in time. In Figure 1, I show an example of a single voicemail greeting clip with measurements of voice frequency in Hz across time. This figure illustrates that the voice frequency (F_0) exhibits significant variation over the duration of the 3-second greeting. It is this variation that is used to estimate the lawyer's voice frequency density function. The mode (or modes) of each lawyer's density function can then be identified to study the possibility that codeswitching manifests itself through lawyers' modulation of their voice frequencies.

However, the exercise of measuring voice frequency is not straightforward. It requires transforming continuous analog data (i.e., the sound produced in original conversations) to discrete digital data (a series of ones and zeros). This "discretization" process introduces measurement error, pursuant to which some secular features of the original sound are sacrificed. In the context of VoIP, a second form of measurement error occurs in cases where there is "packet loss," which refers to a standard Internet protocol that drops bits of the discretized audio data to prevent latency under congested network conditions.[10]

The acoustic data contained in the voicemail greetings were stored on WAV files. I used Praat, a common open-source software program, to measure voice frequency during periods of speech.

[9] The details on how the poor quality was detected are available in Materials and Methods.

[10] I elaborate on these measurement issues in Appendix C.

This was achieved by using a floor of 50 Hz. Coupled with a ceiling of 400 Hz, these bounds helped minimize the interference of sound other than the lawyer's voice, such as background noise.[11]

The sampling rate for each lawyer's voicemail greeting is 8,000 samples per second (every 1/8th of a millisecond, the discretization process yields one sample of the voice frequency) per lawyer.[12] To estimate the shape of each lawyer's voice frequency density function, some experimentation in striking the optimal balance between computational resources and accuracy was required. This process resulted in the extraction of 100 voice frequency quantiles from each lawyer's three-second trimmed voicemail greeting WAV file.

I estimate each lawyer's voice frequency density function as follows: let $h_i(\ )$ be the empirical voice frequency density function based on the voice frequency samples from the WAV file for lawyer i. Then, using Praat, for $q = 1, ..., 100$, each lawyer's voice frequency quantile in Hz, $F_0^{i,q}$, is defined as the highest sample such that:

$$\int^{F_0^q} h_i(\nu)\, d\nu \leq q/100. \tag{2}$$

As a result of this process, the voice frequency quantiles associated with lawyer i are a sequence of 100 numbers $\left\{F_0^{i,1}, F_0^{i,2}, \ldots, F_0^{i,100}\right\}$.[13]

4.3 Classification of Voicemail Greetings

A central challenge in preparing the dataset of lawyers' voice frequencies for analysis was that not all lawyers self-record their voicemail greetings. Voicemail greetings fall into one of two general categories: those recorded personally by the lawyer ("type 1"), and those recorded by someone other than the lawyer ("type 0"). With such a large sample of lawyers, manually determining the greeting type used by a given lawyer in the sample was impractical. Therefore, I used machine learning (ML) techniques to classify the voicemail greetings.[14] This process had a number of steps.

First, I set aside a sub-sample (the "ML sample") of lawyers to train, validate and test a ML model. Overall, the ML sample constituted about 10 percent of the data. The voicemail greetings

[11] I examine other floors and ceilings and present these in the Robustness Checks Subsection.

[12] In practice, the size of each WAV file varies and is smaller than 8000 samples per second of playback time due to "packet loss".

[13] Technically speaking, Praat chooses F_0^q to be the highest frequency sample such that no more than q percent of the samples are less than or equal to F_0^q. In theory, this means sorting by frequency and stopping every 1 percent of the samples (240 for three seconds of audio). One can obtain the mean sample and other moments for each interval of samples between quantiles.

[14] I considered four types of ML models. Details on these ML models and their performance are available in Materials and Methods.

for the lawyers in the ML sample were manually classified as type 1 and type 0 by listening to the content of the voicemail. It was easy to determine whether the lawyer was introducing himself or herself in the greeting or, alternatively, that someone other than the lawyer was speaking.

Second, detailed demographic information on the lawyers was collected from their firm webpages and merged into the ML sample. This information included: title (most commonly, partner or associate), practice area, law school, undergraduate school, any higher degrees (e.g., LL.M., SJD), graduation year for each degree earned, academic honors earned, and gender. Gender was assessed in the following manner: by scrutinizing lawyers' first names, subjectively classifying photos, searching for any mention of pronouns in the biographical description of the lawyer, and cross-referencing the recorded greeting. Generally, these gender indicators perfectly corroborated one another.

Third, the verbal content of each voicemail greeting was analyzed. Using IBM's Watson speech-to-text API, the sentence(s) in the greeting were broken into words. This allowed me to measure the number of words used in the greeting, as well as the lawyer's choice of words (and combinations of words), hoping that there would be systematic differences between type 0 and type 1 recordings that would prove useful in training the ML model.[15]

Fourth, a portion of the ML sample was used to train the ML model in correctly classifying type 1 versus type 0 voicemail greetings. At this point, each lawyer observation had hundreds of variables: name, phone number, firm, firm-specific covariates, the 100-quantile sequence of the lawyer's voice frequency data, other acoustic measurements (i.e., voice intensity), verbal content measures relating to the voicemail greeting, lawyer demographic covariates, and the manually-coded type 1/type 0 voicemail greeting classification. To maximize the performance of the ML model, all available data for each lawyer was employed in the training exercise. While it is difficult to independently identify the contribution of any specific variable to the predictive capacity of the final ML model ("XGBoost") that I used, a ranking of the relative importance of the variables used in XGBoost showed that acoustic measurements played the most significant role.[16] The number of words spoken by the lawyer in the greeting was the most important non-acoustic variable (coming in at sixth place). Gender was the most important–and indeed the only–demographic variable identified in the top-fifteen ranking (coming in at eighth place).

Fifth, the final ML model was tested on the portion of the ML sample that was not used for training and validation. Based on a simple 0.5 probability threshold, the final ML model classified 93.52 percent of the ML test (non-training/validation) sample correctly. In two-thirds of the ML test sample of recordings that met a higher classification threshold (i.e., Prob(type 1) > 0.95 or

[15]For a summary of the most common words used in type 0 and type 1 greetings in the pilot sample, see A2.

[16]The importance of any particular variable/attribute is calculated for a single decision tree by the amount that each attribute split point improves the performance measure, weighted by the number of observations for which the node is responsible. The attribute's importance is then averaged across all decision trees within the model.

Prob(type 1) < 0.05)), the final ML model performed at 99 percent accuracy.

Sixth, the final ML model was used to classify (as type 1 or type 0) the voicemail greetings of the remaining 48,834 lawyers in the dataset (that is, the lawyers that were not part of the ML sample). For 39,967 of these lawyers, demographic information obtained from the ALM Legal Compass database, a leading directory of lawyers, was merged into the dataset by phone number and lawyer name. These 39,967 lawyers, drawn from 86 firms and 696 offices across the U.S., comprise the main sample on which the analysis in the paper is performed. 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).

Table 3 summarizes the number of observations by gender and the two probability thresholds I use. Approximately half of the voicemail greetings (n = 21,405) are classified as type 1 using the low 0.5 threshold; however, prediction errors (i.e., false positives) can be mitigated by using the higher 0.95 threshold at the cost of reducing the type 1 sample size by one-third (n = 14,366).

5 Codeswitching as a Concept

The concept of codeswitching behavior with respect to a malleable attribute like voice frequency is premised on having an in-group and an out-group in a given setting. The in-group and the out-group must have single primary modes of behavior, and these modes must be sufficiently divergent from one another to be observable to the econometrician even in light of measurement error.[17] Evidence that the out-group "switches" to, or utilizes, the in-group's mode of behavior rather than its native out-group mode would be suggestive of pressure to conform to the in-group-driven market norm. I hypothesize that, in the male-dominated elite law firm environments that I study, pressures to conform to workplace norms is evidenced by female lawyers' codeswitching with respect to their voice frequency mode.

To assess whether the premises of codeswitching (two divergent primary modes of the in-group and out-group) are present with respect to male and female voice frequencies, aggregation of the voice frequency data from the groups in question is necessary.[18] To do this, I first remove the threat of confounding lawyer fixed factors, such as the mass and length of one's vocal cords as

[17]For example, consider a symmetric mixture of two normal distributions with equal variances, then the difference between the means must exceed two standard deviations to identify even the slightest bimodality. In practice, the necessary divergence between means is even greater when the mixture is asymmetric. For a recent discussion see Schilling et al. (2002).

[18]Even if a particular lawyer's voice frequency density is bimodal, this alone does not provide evidence for the type of codeswitching this paper aims to establish. The shape of any particular lawyer's voice frequency density may be driven by idiosyncratic factors that have little to do with pressures to conform to in-group norms.

well as other lawyer characteristics, by subtracting the mean voice frequency of each lawyer, $\overline{F_0^i}$, from each of that lawyer's voice frequency quantiles: $F_0^{i,q} - \overline{F_0^i}$. These lawyer-demeaned quantiles are the data used to estimate the voice frequency densities at the group level.

To estimate the voice frequency density of a given group, I use non-parametric kernel density estimation methods. This approach circumvents asymptotic issues that may arise when applying parametric or semi-parametric methods that assume a particular data-generating process of the voice frequency. Further, because my data are selected from a particular non-random snapshot of a lawyer's voice frequency (that is, the voicemail greeting), there are no obvious grounds for assuming stationarity or any stochastic process that governs the voice frequency in general.[19]

Finally, it is important to distinguish between the mechanics of codeswitching and vocal fry, a recent fad among young professional female adults (Anderson et al., 2014). The vocal fry, which typically occurs at the end of a sentence (Wolk et al., 2012), and consists of a train of discrete, low-frequency, glottal pulses produced by the larynx, results in a distinctive acoustic signal utilizing the lowest register in phonation.[20] In contrast, codeswitching entails mixing in-group and out-group voice frequency modes utilizing the modal phonation register–the most frequently used register in speech–significantly above the vocal fry register. Voice frequencies in the modal register average 118 Hz for men and 211 Hz for women, whereas vocal fry averages 50 Hz for both males and females (Blomgren et al., 1998). To sum, vocal fry is a significantly lower phonation register than the one I study, and is more likely to occur at the end of a sentence. In the Robustness Checks, I show both voice frequency over time and with varying frequency floors to rule out the possibility that vocal fry is confounding my results.

6 Results

I offer evidence that the premises of the codeswitching concept are reasonable in the setting that I study: males and females in an elite law firm. To assess whether males have a primary voice frequency mode, I use type 1 (self-recorded) voicemail greetings of male lawyers. To assess whether females have a primary voice frequency mode, I exploit the fact that type 0 voicemail greetings (recorded by an assistant, introducing the lawyer in the third person) are recorded by females. This allows me to explore the codeswitching concept by studying female assistants' voices (that is, female non-lawyers) as a proxy for the female primary mode. This assists in assessing whether female lawyers–in the specific setting of male-dominated elite law firms as well as in the particular

[19]I follow the direct local time density estimation argument of Wang and Phillips (2009), which make the estimation approach of stochastically nonstationary time series data closely related to conventional nonparametric estimation.

[20]The physiological properties of vocal fry are characterized by a glottal area function that has sharp, short pulses followed by a long closed glottal interval. There is a very short open period and a very long period where the vocal folds are completely adducted.

context of introducing themselves as lawyers–stick to the female primary mode. If, alternatively, they choose voice frequency behavior that is distinct from the female primary mode exhibited by female assistants and instead mimics the voice frequency of the primary male mode, such behavior is consistent with the codeswitching concept.

To maximize the accuracy of my results, I use the voicemail greetings that meet the high-threshold classification (i.e., Prob(type 1) > 0.95 and Prob(type 0) > 0.95). Figure 2 plots kernel density estimates for each of these two types of recordings. To compare the densities on a common horizontal axis, the relevant group mean of each type, $1/g \sum_{i \in g} \overline{F_0^i}$, is added to the group's density. As seen, the voice frequency densities in both panels are unimodal and symmetric around their primary modes. In the lefthand-side figure, the voice frequency density of male lawyers is highly concentrated around 100 Hz. This is the only observable mode in the data, and suggests that there is a primary male mode of voice frequency. Likewise, in the righthand-side figure, the voice frequency density of female assistants is also unimodal, and centered around 200 Hz. This primary female mode of voice frequency represents a significant divergence from the male mode that is easily observable.[21]

6.1 Key Findings

Figure 3 presents the key findings using both the low- and high-thresholds for classifying type 1 voicemail greetings. Beginning with the former, the righthand-side figure shows the kernel density estimates for both female and male lawyers. The figure indicates that male lawyers have a primary (and only) voice frequency mode around 100 Hz and female lawyers have a primary voice frequency mode around 200 Hz, similar to the "female mode" identified in the female assistants above. However, unlike the female assistants, there is a discernible secondary voice frequency mode for the female lawyers at 100 Hz, which overlaps precisely with the male mode. Further, in the lefthand side figure, produced from the smaller yet more accurately classified sample of voicemail greetings, this secondary mode is easier to detect. No such secondary modes are observed in the male lawyer densities, nor are any additional modes observable in the female assistant densities. These results constitute the main evidence in support of codeswitching behavior: the key distinguishing feature of codeswitching behavior is the ability to preserve the integrity of each underlying norm and the prescribed conventions associated with it. Specifically, it is essential that both the male and female modes are utilized by the lawyer. That codeswitching behavior is

[21] To check for robustness, I also analyze the pilot sample of data that was manually classified by a human listener: a set of voicemail greetings recorded by male lawyers, and a set voicemail greetings recorded entirely in third person by females (i.e., without an interruption by the lawyer to introduce him or herself by name). The findings using this sample are consistent with those of the larger sample. Estimation results using the pilot sample are available in Figure A3.

detected among female but not male lawyers is consistent with the hypothesis that pressures to conform to norms are disproportionately borne by the out-group.

6.2 Lawyer Status and Codeswitching

The status of a worker within the out-group can also influence her degree of codeswitching. Specifically, given that associates are in a precarious position (i.e., more marginalized as compared to partners), pressures to conform to in-group norms may be stronger for out-group associates relative to out-group partners. To the extent that these pressures translate into a greater degree of conformity to in-group norms, one might expect female associates to utilize the male mode more than female partners.

Using the high-threshold classification of voicemail greetings, Figure 4 presents estimation results of voice frequency densities for type 1 associates and partners separately. Consistent with the hypothesis, codeswitching behavior is more pronounced among female associates than among female partners. Specifically, the secondary mode is significantly more discernible in the voice frequency density of female associates relative to female partners. This suggests that, conditional on making partner, female lawyers may feel less pressures to conform to in-group norms because of critical gains in job security, status, or pay that typically follow a promotion from associate to partner.

The results above suggest that other dimensions of within-group heterogeneity in codeswitching behavior may exist, where pressures to conform may not be experienced by all members equally. One dimension that has a direct effect on these pressures is one's natural voice pitch as measured by the mean voice frequency of the lawyer. In particular, I conjecture that if pressures to conform drive codeswitching behavior, then female lawyers with a high voice pitch must exert greater switching effort to utilize the male mode than female lawyers with a low voice pitch. To test this hypothesis, I compare voice frequency densities of lawyers with above-median mean voice frequency ("high") to lawyers with below-median mean voice frequency ("low"). If the distance between modes is greater among lawyers with a high voice than lawyers with a low voice, then it would suggest fine-tuning of the voice frequency by the lawyers to reach a specific target. I present evidence for this hypothesis in Figure 5 using the high classification threshold sample of type 1 voicemail greetings. The lefthand-side plot shows that female lawyers with a low voice have a distinctive secondary mode close to their primary mode, with a short left tail. Relative to these lawyers, the righthand-side plot shows that female lawyers with a high voice have a flatter secondary mode with a peak farther away from their primary mode and a long left tail. These findings suggest differential pressures to conform by female lawyers to the male voice frequency mode.

I explore other dimensions of heterogeneity in codeswitching behavior among female lawyers, including practice area (e.g., litigators versus non-litigators), female group size, and share of female partners at both the firm level and the office level (see Figures A4, A5 and A6 in Appendix A). Further, because culture and attitudes toward norms can depend on firm history, I examine differences in firms established after versus before 1917, the median year in the sample.[22] Overall, I do not find evidence for heterogeneity in voice frequency densities among female lawyers along these firm characteristics. There are at least two explanations for this: (1) the firms in my study are highly homogenous with respect to workplace norms, or (2) the identity one chooses to convey on a voicemail greeting is based on market norms rather than firm-specific norms.

6.3 Robustness Checks

To corroborate my main findings, I use a representative sample of 2,255 recordings I used to pilot my methodology. This sample serves two main purposes in this section. First, I examine whether my results are robust to an alternative classification of voicemail greetings. Specifically, I reclassify a formerly type 0 voicemail greeting as type 2 if it contains a brief personal identification of the lawyer by name. Similar to the main sample, about one half of the sample are classified as type 1. However, about two third of non-type 1 recordings are now reclassified as type 2 ('mixed'). See Table A1 for a summary of this three-way classification.

Using the same empirical strategy as before (i.e., 100 lawyer-demeaned voice frequency quantiles per recording), Figure 6 presents kernel density estimation results for each of the three types of manually-classified voicemail greetings. In the middle panel, type 1 densities replicate the main findings, where the male density is unimodal with a mode at 100 Hz, and the female density is bimodal with a primary mode at 200 Hz and a secondary mode at 100 Hz. On the lefthand-side, the manually classified type 0 densities of female and male lawyers are remarkably similar: the densities are nearly on top of each other, and resemble a normal distribution with mean 200 Hz. On the other hand, the newly-classified type 2 voice frequency densities are unimodal. However, relative to a voicemail greeting recorded in the third person for a female lawyer, the inclusion of a male voice (upon self-identifying) in a third-person recording for a male lawyer shifts the density downward 10 to 15 Hz.[23] These results corroborate the conjecture that the location shift between males and females observed is due to voicemail greetings that include a brief self-identification by the lawyer.

I next use this sample to examine whether the main results are sensitive to the duration of the

[22]Likewise, I used LLM and LLB degrees to proxy for lawyers in these firms who did not grow up in the US, but there are few and these few do not speak differently from lawyers who went to college and law school in the US.

[23]Consistent with these findings, the densities of type 0 male lawyers in the full sample are slightly shifted to the left of type 0 female lawyers. Figure A7 presents kernel density estimation results of type 0 lawyers by gender.

voicemail recording. In the analysis above, I use only the first three seconds of each voicemail greeting. Most voicemail greetings conclude within five seconds. The voicemail recordings were clipped to three seconds primarily because many type 1 greetings switch to a third person toward the end of the recording to provide instructions on how to proceed with leaving a message. On the other hand, a clip that is too short may not contain sufficient audio data for analysis. In Figure 7, I provide kernel density estimation results using data from the first 1, 3 and, 5 seconds of each voicemail greeting. The secondary mode of female lawyers is undetectable using only the first second of the voicemail greeting. Overall, the results suggest that the secondary mode of female lawyers is discernible as more time lapses from the beginning of the voicemail greeting.[24]

Finally, I estimate asymptotic confidence intervals for the voice frequency densities of female lawyers following Fiorio (2004). I find that the densities are precisely estimated with confidence intervals nearly overlapping with the density estimates. Likewise, I implement a variable bandwidth approach described in Salgado-Ugarte et al. (2003), where bandwidth length is wider in low density areas and narrower in high density areas. This produces observation-specific bandwidths and addresses a potential concern that voice patterns are driven by random differences between voice frequency quantiles.[25] Estimates based on the variable bandwidth algorithm confirm the key findings. At the same time, several approaches for testing whether the true unobservable population density has a specific number of modes are based on a theory developed by Silverman (1981).[26] However, visual inspection of non-parametric density estimates with varying bandwidths remains the dominant form of carrying out robustness checks and evaluating the evidence on the shape of a density.

7 Between-Lawyer Voice Frequency Analysis

To provide context for my findings on codeswitching behavior, I describe the cross-sectional variation of lawyers in this sector. For this exercise, I use the mean voice frequency in each voicemail greetings ($\bar{F}_0^i$) as representative of a lawyer's voice frequency.

[24] Relatedly, I find voice frequency for men and women to evolve in parallel. Using a similar methodology to the quantile selection, Praat can sequentially select the voice frequency sample closest to satisfying equation 2 for each 0.03 seconds of playback. See Figure B1.

[25] Visually, this method is analogous to binned scatterplots, where an estimate is calculated for each fixed quantile rather than a range of the data.

[26] 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 $k = 1, 2, \ldots$ the smallest bandwidth that can support k modes or less and (2) generating smoothed bootstrapped samples from each critical density using a Gaussian kernel. The proportion of samples with more modes than k reflects the significance of the bandwidth cutoff. The test is seen as conservative since the bootstrap samples are drawn from the critical density only. Put differently, if the true density is bimodal, it does not follow that the data are drawn only from the *most* bimodal distribution.

7.1 Male Lawyers versus Female Lawyers

Following a similar approach to that used for assessing codeswitching behavior, I estimate kernel densities of between-lawyer mean voice frequencies. Figure 8 shows that the mean voice frequency density of type 1 females is bimodal with a primary mode at 200 Hz and a minor mode at the 100 Hz range overlapping with the primary (and only) male mode. Further, the secondary female mode is more pronounced using the high classification threshold (Pr(type 1) > 0.95) relative to the low classification threshold (Pr(type 1) > 0.5). As in the previous results, this is likely because a greater proportion of voicemail greetings are misclassified as type 1 in the low-threshold sample. This misclassification hampers the ability to detect secondary modes.

Analyses of these data suggest that the bimodality in mean voice frequency among female lawyers is insensitive to the inclusion of covariates. Figure 9 illustrates results from sensitivity tests on type 1 lawyers. In this figure, I compare the raw distribution of each gender group presented in the previous figure (Figure 8) to the distribution of residuals obtained from regressing the mean voice frequency per lawyer on a set of title, city, and firm fixed effects. The distributions essentially sit on top of each other for both male and female lawyers, respectively.[27] That said, unlike the (within-lawyer) results on codeswitching behavior, I cannot rule out the possibility that the distribution of an omitted variable confounds these findings.

In contrast to voicemail greetings recorded by female lawyers, the mean voice frequency density of type 0 lawyers has only one detectable mode. Similar to the within-lawyer results, the density of type 0 male lawyers is slightly shifted to the left of type 0 female densities due to mixed (type 2) greetings. Overall, these findings suggest that a small, yet discernible, fraction of female lawyers share a common mean voice frequency with male lawyers.[28]

7.2 Associates versus Partners

Because prospects for promotion in a male-dominated sector may depend on one's voice pitch, I next examine whether female partners are more likely to populate the secondary mode than female associates, relative to the primary mode. Put differently, I ask whether the data support the hypothesis that female lawyers with a "male" mean voice frequency are more likely to be promoted to partner relative to female lawyers with a "female" mean voice frequency.

I first show a scatter plot of the mean voice frequencies in my sample of type 1 associates and partners across firms using the high classification threshold. Figure 10 presents the data sorted by firm rank. First, there appears to be no discernible differences in voice frequencies among male

[27] Figure A9 provides similar results using a pilot sample of manually-classified type 1 lawyers with a rich set of lawyer characteristics and fixed effects.

[28] Figures A10 and A11 provide further evidence that this group of female lawyers is pervasive across experience levels and firms, respectively.

lawyers by title (partner versus associate) or by firm rank. By and large, the mean voice frequency of male lawyers is clustered in a relatively narrow band centered around 100 Hz. In contrast, there are two noticeable differences between female associates and female partners: (1) the primary mode of female partners is lower than the primary mode of female associates by 10-15 Hz, and (2) female partners are significantly more represented in the secondary mode across firms than female associates. Only a differential change in the shape of the densities would support the hypothesis that female lawyers in the secondary ("male") mode have a higher promotion rate than female lawyers in the primary ("female") mode. Figure 11 plots the kernel density estimates separately for partners and associates. It can been seen that the female partners' density has more mass at the "male" mode relative to the female associates' density. This suggests the female lawyers who utilize the male mode are more likely to be promoted to partner than female lawyers who utilize the female mode.

7.3 Unobserved Heterogeneity

To provide further structure to the cross-sectional results, I apply a methodology developed in Deb and Trivedi (1997) and Deb (2012) to estimate a model of unobserved heterogeneity with two latent groups of female lawyers. Overall, the model supports the hypothesis that the data on mean voice frequencies of type 1 female lawyers are generated by a mixture of two distributions: one group ('Group 1') accounts for about 7 percent of the population with a mean voice frequency of 140 Hz, whereas the other group's ('Group 2') mean voice frequency is 200 Hz.

Most interesting with respect to this methodology is the difference between the estimated groups among female associates and female partners. In Table 4, I present the results of estimating the model separately for female associates and female partners with firm fixed effects. Figure 12 shows the estimated histograms of each latent group of female lawyers within a firm. As seen, the overlying kernel density estimates understate the degree of separation of the mean voice frequencies between both groups of female lawyers. In particular, the leftward shift seen for partners relative to associates among Group 1 lawyers is 30 Hz, twice the estimated shift observed in Group 2. This difference is partially explained by the relatively higher compression among Group 1's partners relative to associates. There is also significantly less overlap in mean voice frequencies between both groups of partners relative to associates. Unlike the density of Group 2 lawyers, which exhibits narrow and symmetric changes in variation, the density of Group 1 lawyers changes asymmetrically from associate to partner. Together, results from this model suggest that, within a given firm, segregation in mean voice frequency is higher among female partners than among female associates. In terms of estimated shares, Group 1 is 6.3 and 7.6 percent of the population of associates and partners, respectively. This difference is statistically insignificant, but implies that Group 1 female lawyers, characterized by a relatively low mean voice frequency, may have an advantage over Group 2 female lawyers.

Finally, one important dimension of voice that I have not focused on is amplitude. In a separate exercise, I use both voice amplitude and frequency to examine whether this richer set of data can better predict lawyer observables than data on voice frequency alone. To do this, I utilize an unsupervised machine learning method, where the model iteratively searches for the optimal classification of lawyers into clusters without any additional information on the lawyers.[29] See Appendix D for more details. While there is some evidence that voice amplitude may contribute to the model's performance in predicting lawyer attributes, amplitude data in most cases amount to introducing noise to the prediction process suggesting that a better understanding of the relationship between both data dimensions is required. I leave this as an area for future research.

8 Conclusion

In the first large-scale study of voice in the labor market, I document a bimodal distribution density function of voice frequency for female lawyers: a primary female mode of voice frequency at about 200 Hz as well as a secondary female mode at about 100 Hz that is coextensive with the primary male voice frequency mode. Likewise, consistent with codeswitching and their lower status in their firms' hierarchies, female associates are more likely than female partners to switch between the primary female voice frequency mode and the primary male mode. Finally, differences between the distributions of mean voice frequencies for female associates and female partners suggest that female lawyers that utilize the male mode are slightly more successful at making partner.

These findings are consistent with the idea that pressures to conform to in-group norms are borne disproportionately by marginalized workers. Data from a cross-section of studies suggest that female voice frequencies have dropped by over 20 Hz over the latter half of the twentieth century (Pemberton et al., 1998). However, in-group norms can change over time and thus make it impossible for out-group members to meet the standard (Pesendorfer, 1995). Such punitive dynamics for females in the labor market is presented in Kuziemko et al. (2018), and recent evidence on the role of voice in the labor market suggests an ever-changing fine line that female workers are required to walk:

> *Women seem to be damned whatever they do. Speak loudly and they are deemed shrill; speak softly and they are meek. A high voice is unserious. Low-frequency vocal fry is off-limits too.*
>
> –"Women's voices", *The Economist*, October 4, 2018

[29] I thank Sendhil Mullainathan for suggesting this approach including the application of the Fourier Transform to my data.

References

Akerlof, G. A. (1980). A theory of social custom, of which unemployment may be one consequence. *Quarterly Journal of Economics 94*(4), 749–775. 5

Akerlof, G. A. and R. E. Kranton (2000). Economics and identity. *Quarterly Journal of Economics 115*(3), 715–753. 2

Anderson, R. C., C. A. Klofstad, W. J. Mayew, and M. Venkatachalam (2014). Vocal fry may undermine the success of young women in the labor market. *PloS one 9*(5), e97506. 13

Apple, W., L. A. Streeter, and R. M. Krauss (1979). Effects of pitch and speech rate on personal attributions. *Journal of Personality and Social Psychology 37*(5), 715. 7

Argyle, M. (1972). Non-verbal communication in human social interaction. 6

Auer, P. (2013). *Code-switching in conversation: Language, interaction and identity*. Routledge. 6

Austen-Smith, D. and R. G. Fryer (2005). An economic analysis of 'acting white'. *Quarterly Journal of Economics*, 551–583. 5

Banse, R. and K. R. Scherer (1996). Acoustic profiles in vocal emotion expression. *Journal of personality and social psychology 70*(3), 614. 6

Baus, C., P. McAleer, K. Marcoux, P. Belin, and A. Costa (2019). Forming social impressions from voices in native and foreign languages. *Scientific Reports 9*(1), 414. 6

Benjamin, D. J., J. J. Choi, A. J. Strickland, et al. (2010). Social identity and preferences. *American Economic Review 100*(4), 1913–1928. 5

Bernheim, B. D. (1994). A theory of conformity. *Journal of Political Economy 102*(5), 841–877. 5

Biddle, J. E. and D. S. Hamermesh (1998). Beauty, productivity, and discrimination: Lawyers' looks and lucre. *Journal of Labor Economics 16*(1), 172–201. 7

Blomgren, M., Y. Chen, M. L. Ng, and H. R. Gilbert (1998). Acoustic, aerodynamic, physiologic, and perceptual properties of modal and vocal fry registers. *The Journal of the Acoustical Society of America 103*(5), 2649–2658. 13

Buller, D., J. Burgoon, and W. Woodall (1996). *Nonverbal communications: The unspoken dialogue*. New York: McGraw-Hill. 7

Bursztyn, L., T. Fujiwara, and A. Pallais (2017). 'acting wife': Marriage market incentives and labor market investments. *American Economic Review 107*(11), 3288–3319. 5

Coate, S. and G. C. Loury (1993). Will affirmative-action policies eliminate negative stereotypes? *American Economic Review*, 1220–1240. 5

Cornell, B. and I. Welch (1996). Culture, information, and screening discrimination. *Journal of Political Economy 104*(3), 542–571. 5

Das, J., S. B. De Loach, et al. (2011). Mirror, mirror on the wall: The effect of time spent grooming on earnings. *Journal of Behavioral and Experimental Economics 40*(1), 26–34. 7

Deb, P. (2012). Fmm: Stata module to estimate finite mixture models. 19

Deb, P. and P. K. Trivedi (1997). Demand for medical care by the elderly: A finite mixture approach. *Journal of applied Econometrics 12*(3), 313–336. 19

Ekman, P., W. V. Friesen, and K. R. Scherer (1976). Body movement and voice pitch in deceptive interaction. *Semiotica 16*(1), 23–28. 7

Fang, H. (2001). Social culture and economic performance. *American Economic Review*, 924–937. 5

Fiorio, C. (2004). Confidence intervals for kernel density estimation. *Stata Journal 4*(2), 168–179. 17

Fryer, R. G. and P. Torelli (2010). An empirical analysis of acting white. *Journal of Public Economics 94*, 380–396. 5

Gardner-Chloros, P. (2009). *Code-switching*. Cambridge University Press. 5

Giles, H. and P. Powesland (1997). *Accommodation theory*. Springer. 6

Grogger, J. (2011). Speech patterns and racial wage inequality. *Journal of Human Resources 46*(1), 1–25. 6

Heller, M. (1988). *Codeswitching: Anthropological and sociolinguistic perspectives*, Volume 48. Walter de Gruyter. 4

Heller, M. (1992). The politics of codeswitching and language choice. *Journal of Multilingual & Multicultural Development 13*(1-2), 123–142. 6

Imhof, M. (2010). Listening to voices and judging people. *Intl. Journal of Listening 24*(1), 19–33. 7

Jeffries, M. P., T. L. Gosa, and E. Nielson (2015). The king's english: Obama, jay z, and the science of code switching. *The Hip Hop & Obama Reader*, 243–61. 4

Klatt, D. H. (1973). Discrimination of fundamental frequency contours in synthetic speech: Implications for models of pitch perception. *The Journal of the Acoustical Society of America 53*(1), 8–16. 8

Klofstad, C. A., R. C. Anderson, and S. Peters (2012). Sounds like a winner: Voice pitch influences perception of leadership capacity in both men and women. *Proceedings of the Royal Society of London B: Biological Sciences 279*(1738), 2698–2704. 7

Kollmeier, B., T. Brand, and B. Meyer (2008). Perception of speech and sound. In *Springer handbook of speech processing*, pp. 61–82. Springer. 8

Kuziemko, I., J. Pan, J. Shen, and E. Washington (2018). The mommy effect: Do women anticipate the employment effects of motherhood? Technical report, National Bureau of Economic Research. 20

Lang, K. (1986). A language theory of discrimination. *Quarterly Journal of Economics 101*(2), 363–382. 5

Lee, S. and K. Ryu (2012). Plastic surgery: Investment in human capital or consumption? *Journal of Human Capital 6*(3), 224–250. 7

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, Mass.: MIT Press. 8

Leongomez, J., M. David, V. R. Little, A. C., and R. S. Craig (2017, 06). Perceived differences in social status between speaker and listener affect the speaker's vocal characteristics. *PLOS ONE 12*(6), 1–21. 6

Lippi-Green, R. (2012). *English with an accent: Language, ideology and discrimination in the United States*. Routledge. 6

Mayew, W. J., C. Parsons, and M. Venkatachalam (2013). Voice pitch and the labor market success of male chief executive officers. *Evolution & Human Behavior 34*(4), 243–248. 6

Mersenne, M. (1636). *Traité de l'Harmonie Universelle* (First ed.). Paris: Pierre Ballard. 7

Myers-Scotton, C. (1995). *Social motivations for codeswitching: Evidence from Africa*. Oxford University Press. 6

O'Hair, D. and M. J. Cody (1987). Machiavellian beliefs and social influence. *Western Journal of Communication (includes Communication Reports) 51*(3), 279–303. 7

Ohala, J. J. et al. (1984). An ethological perspective on common cross-language utilization of f0 of voice. *Phonetica 41*(1), 1–16. 6

Pemberton, C., P. McCormack, and A. Russell (1998). Have women's voices lowered across time? a cross sectional study of australian women's voices. *Journal of Voice 12*(2), 208–213. 20

Pesendorfer, W. (1995). Design innovation and fashion cycles. *American Economic Review*, 771–792. 5, 20

Pisanski, K., E. C. Mora, A. Pisanski, D. Reby, P. Sorokowski, T. Frackowiak, and D. R. Feinberg (2016, 09). Volitional exaggeration of body size through fundamental and formant frequency modulation in humans. *Nature 6*, 34389 EP –. 6

Salgado-Ugarte, I. H., M. A. Perez-Hernandez, et al. (2003). Exploring the use of variable bandwidth kernel density estimators. *Stata Journal 3*(2), 133–147. 17

Schilling, M. F., A. E. Watkins, and W. Watkins (2002). Is human height bimodal? *American Statistician 56*(3), 223–229. 12

Silverman, B. W. (1981). Using kernel density estimates to investigate multimodality. *Journal of the Royal Statistical Society. Series B (Methodological)*, 97–99. 17

Smith, D. R. and R. D. Patterson (2005). The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age a. *Journal of the Acoustical Society of America 118*(5), 3177–3186. 6

Spence, C. (2011). Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics 73*(4), 971–995. 6

Tigue, C. C., D. J. Borak, J. J. O'Connor, C. Schandl, and D. R. Feinberg (2012). Voice pitch influences voting behavior. *Evolution and Human Behavior 33*, 210–216. 7

Vogt, H. (1954). Language contacts. *Word 10*(2-3), 365–374. 5

Wang, Q. and P. C. Phillips (2009). Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory 25*(3), 710–738. 13

Weston, P. S., M. D. Hunter, D. S. Sokhi, I. D. Wilkinson, and P. W. Woodruff (2015). Discrimination of voice gender in the human auditory cortex. *NeuroImage 105*, 208–214. 6

Wolk, L., N. B. Abdelli-Beruh, and D. Slavin (2012). Habitual use of vocal fry in young adult female speakers. *Journal of Voice 26*(3), e111–e116. 13

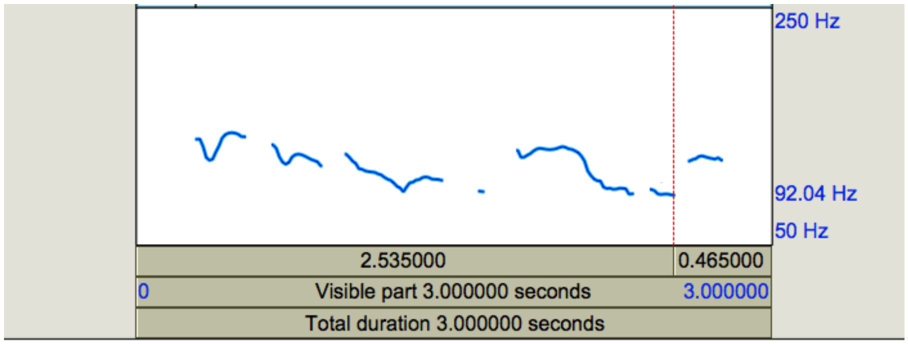

Figure 1: Voice Frequency Variation in Voicemail Greetings

Notes: This figure illustrates the variation in voice frequency (y-axis) over time (x-axis) for a random lawyer in the sample. The audio data is based on a WAV file recording of the first three seconds of the lawyer's voicemail greeting. The blue line marks the voice frequency at every 1/8 of a millisecond (i.e., the sample rate). The blank intervals reflect periods of silence (below 50 Hz) or abnormally high frequencies (above 400 Hz) resulting from background noise. For example, the minimum voice frequency of 92.04 Hz (marked by the red vertical dashed line) is reached at 2.535 seconds from the beginning of the voicemail greeting. The variation in voice frequency over this three-second time interval forms the basis of my data analysis on codeswitching.

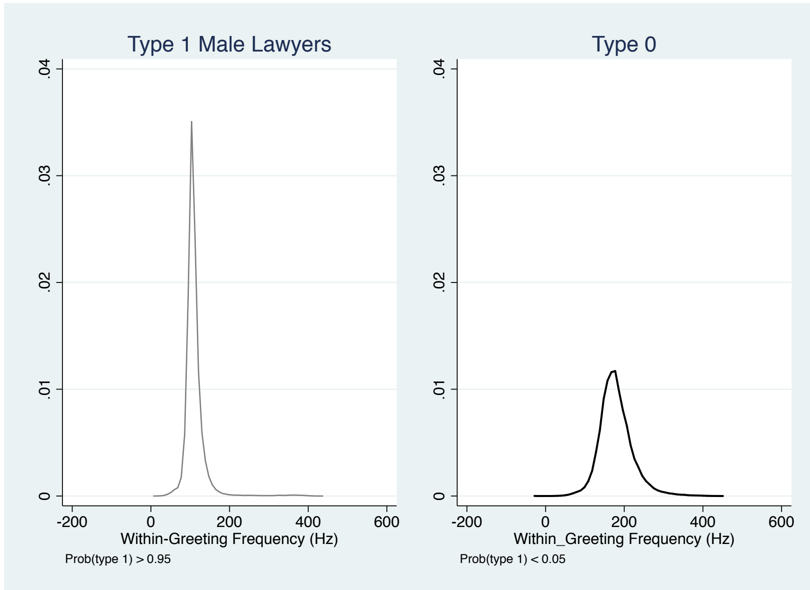

Figure 2: Primary Voice Frequency Modes of Males and Females

Notes: This figure presents kernel density estimates of voice frequencies using the sample of type 1 (self-recorded) and type 0 (recorded in third person) voicemail greetings that meet the high-threshold classification (i.e., Prob(type 1) > 0.95 and Prob(type 0) > 0.95, respectively). The data I use to estimate the densities are 100 greeting-demeaned voice frequency quantiles per voicemail greeting, which are obtained from the first three seconds of each greeting. The lefthand-side figure uses voicemail greetings recorded by male lawyers (n = 10,651), whereas the righthand-side figure uses voicemail greetings recorded in third person by females (n = 12,616).

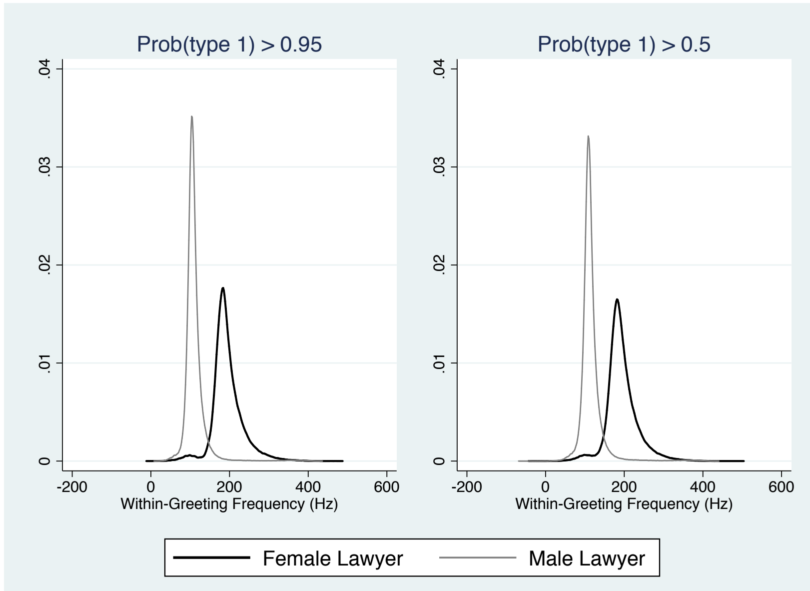

Figure 3: Baseline Results by Predicted Type 1 Thresholds

Notes: This figure presents kernel density estimates of male and female voice frequencies of lawyers using two samples: on the lefthand-side, type 1 voicemail greetings that meet the high-threshold classification (i.e., Prob(type 1) > 0.95), and, on the righthand-side, type 1 voicemail greetings that meet the low-threshold classification (i.e., Prob(type 1) > 0.5). The data I use to estimate the densities are 100 greeting-demeaned voice frequency quantiles per voicemail greeting, which are obtained from the first three seconds of each greeting. The lefthand-side figure uses 14,366 voicemail greetings, whereas the righthand-side figure uses 21,405 voicemail greetings. Voicemail greetings of female lawyers constitute approximately one third of each sample.

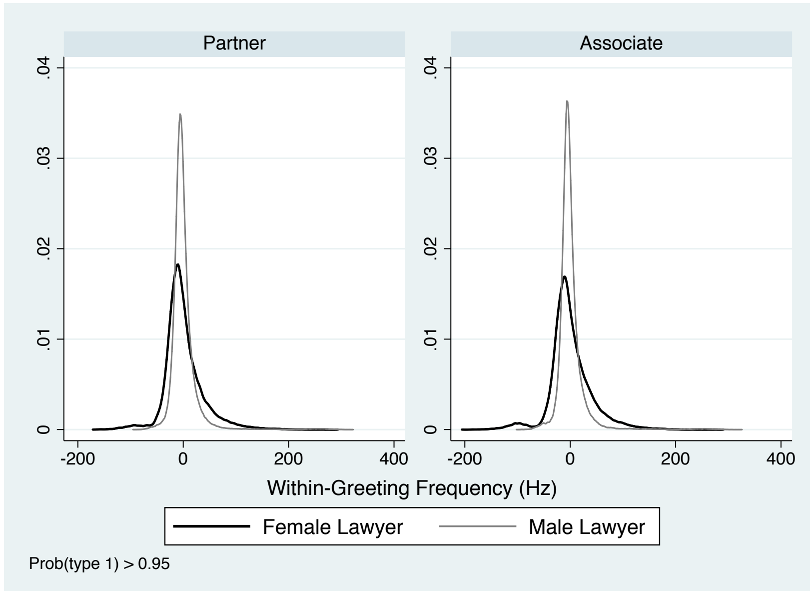

Figure 4: Lawyer Densities by Title

Notes: This figure presents kernel density estimates of voice frequencies by lawyer title and gender. The data used to estimate the densities are type 1 voicemail greetings that meet the high-threshold classification (i.e., Prob(type 1) > 0.95). The data I use to estimate the densities are 100 greeting-demeaned voice frequency quantiles per voicemail greeting, which are obtained from the first three seconds of each greeting.

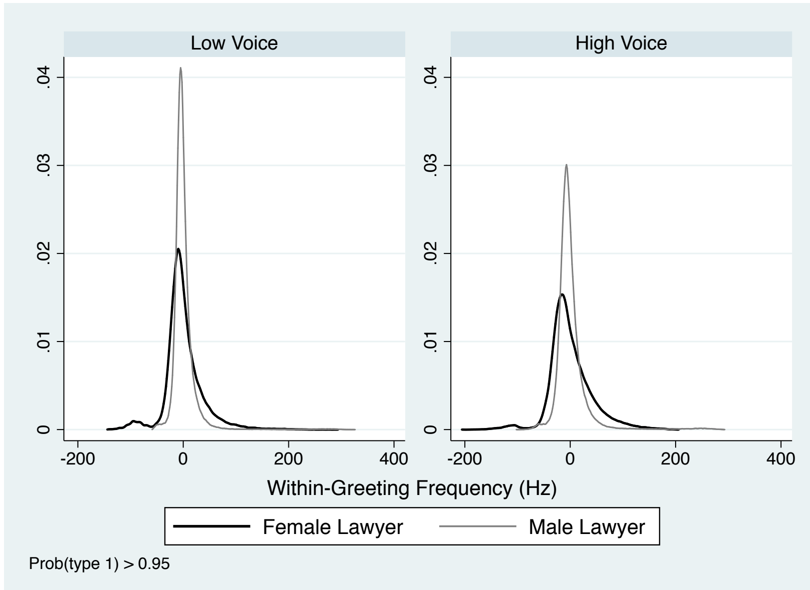

Figure 5: Lawyer Densities by Mean Voice Frequency

Notes: This figure presents kernel density estimates of voice frequencies by mean voice frequency and gender. *Low Voice* is defined as a lawyer with a mean voice frequency below the group median, whereas *High Voice* is defined as a lawyer with a mean voice frequency above the group median. The data used to estimate the densities are type 1 voicemail greetings that meet the high-threshold classification (i.e., Prob(type 1) > 0.95). The data I use to estimate the densities are 100 greeting-demeaned voice frequency quantiles per voicemail greeting, which are obtained from the first three seconds of each greeting..

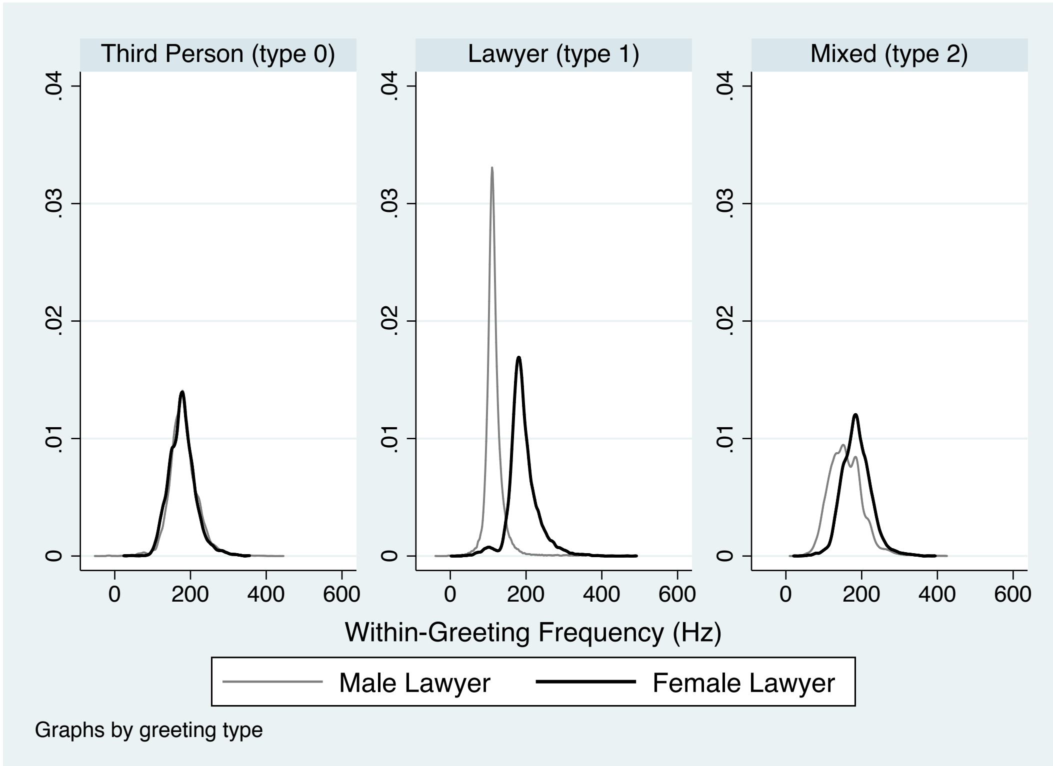

Figure 6: Sensitivity of Lawyer Densities to Voicemail Greeting Classification

Notes: This figure presents kernel densities of voice frequencies based on the pilot sample of voicemail greetings. These data were manually classified into three voicemail greeting types: Type 0 are recorded in third person by females, type 1 are recorded by lawyers, and type 2 are recorded in third person by females with a brief potion of the recording made by the lawyer introducing him or herself by name. The number of voicemails of each type are presented in Table A1. To produce this figure, I used 100 greeting-demeaned frequency quantiles from each voicemail greeting, which are obtained from the first three seconds of each greeting. The average voice frequency of each gender group by type was added to these lawyer-demeaned quantiles.

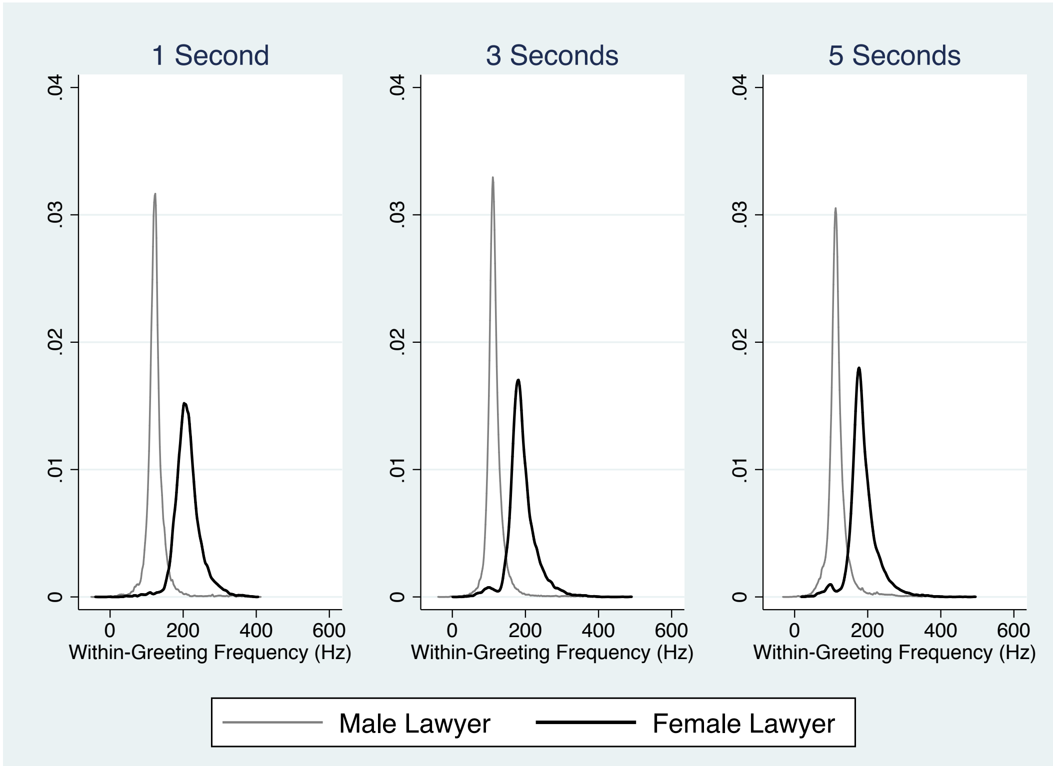

Figure 7: Sensitivity of Lawyer Densities to Voicemail Greeting Length

Notes: This figure presents kernel densities of voice frequencies based on a pilot sample of 1,177 manually-classified type 1 voicemail greetings. To produce this figure, the recordings were first uniformly trimmed to different time lengths: 1, 3, and 5 seconds that lapsed from the beginning of the voicemail greeting. Next, I used 100 greeting-demeaned frequency quantiles from each voicemail greeting. Three samples of data were produced from each original voicemail greeting: The lefthand-side figure uses the sample of voice frequency quantiles obtained from 1 second-long recordings, the middle figure uses the baseline sample obtained from 3 second-long recordings, and the righthand-side figure uses voice frequency data obtained from 5 second-long recordings.

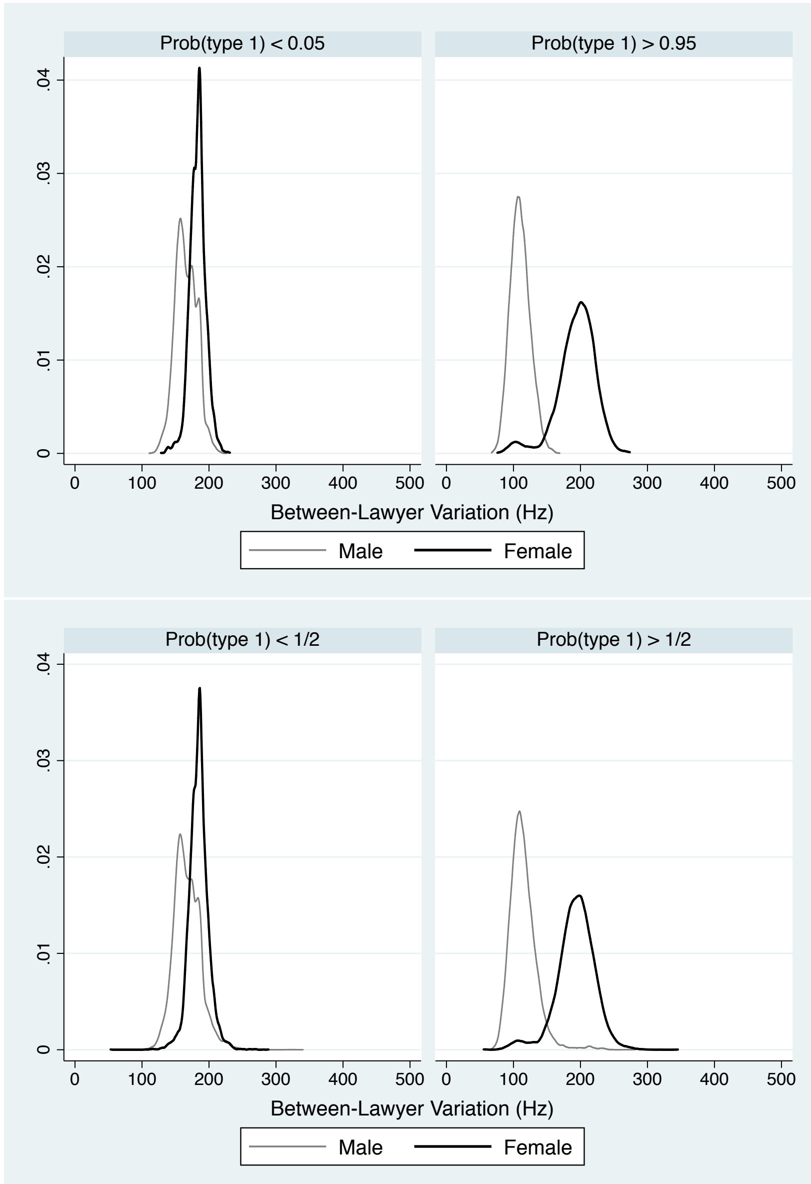

Figure 8: Between-Lawyer Densities by Predicted Greeting Type

Notes: These figures present between-lawyer kernel density estimates of mean voice frequencies. The top figures present results using the high-threshold classification, whereas the bottom figures present results using the low-threshold classification. Likewise, the righthand-side figures present results on type 1 voicemail greetings, whereas the lefthand-side figures present results on type 0 voicemail greetings. The data used to estimate the densities are the mean voice frequency per voicemail greeting, which is obtained from the first three seconds of the greeting. The number of lawyers of each type are presented in Table 3.

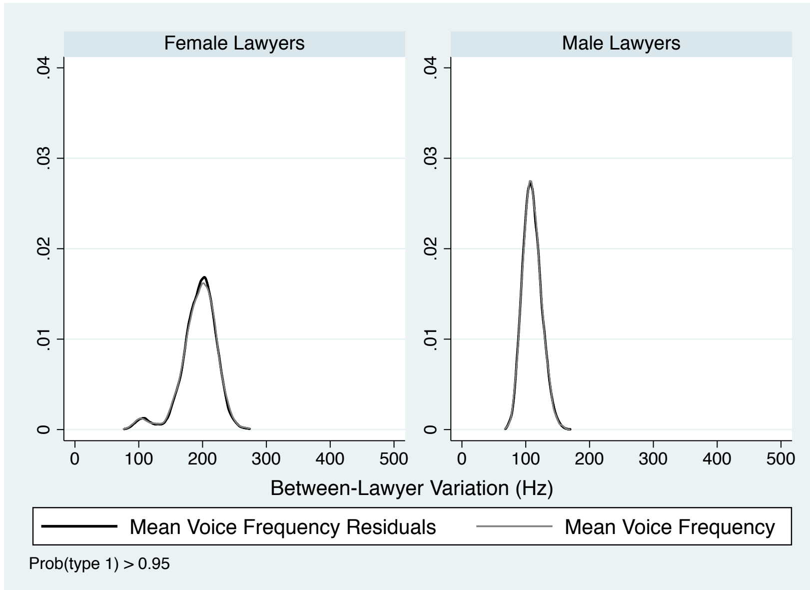

Figure 9: Sensitivity of Between-Lawyer Densities to Controls

Notes: This figure presents kernel density estimates of mean voice frequencies using the sample of type 1 recordings that meet the high-threshold classification. The data used to estimate the densities are the mean voice frequency per voicemail greeting, which is obtained from the first three seconds of the greeting. Residuals are obtained from regressing mean voice frequencies on lawyer city, firm and title fixed effects. The righthand-side figure uses voicemail greetings recorded by male lawyers (n = 10,651), whereas the lefthand-side figure uses voicemail greetings recorded by female lawyers (n = 3,715).

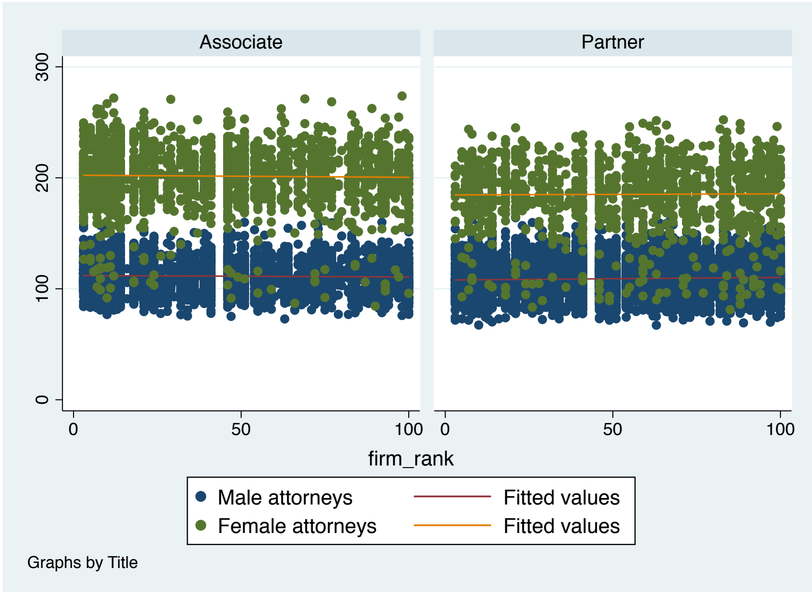

Figure 10: Between-Lawyer Scatter Plots by Firm and Title

Notes: This figure presents scatter plots of mean voice frequencies by title and gender using type 1 voicemail greetings that meet the high-threshold classification. The mean voice frequency per voicemail greeting, obtained from the first three seconds of the greeting, is plotted against the lawyer's firm rank. The solid lines represent the linear fit between mean voice frequency and firm rank for each gender group of lawyers by title.

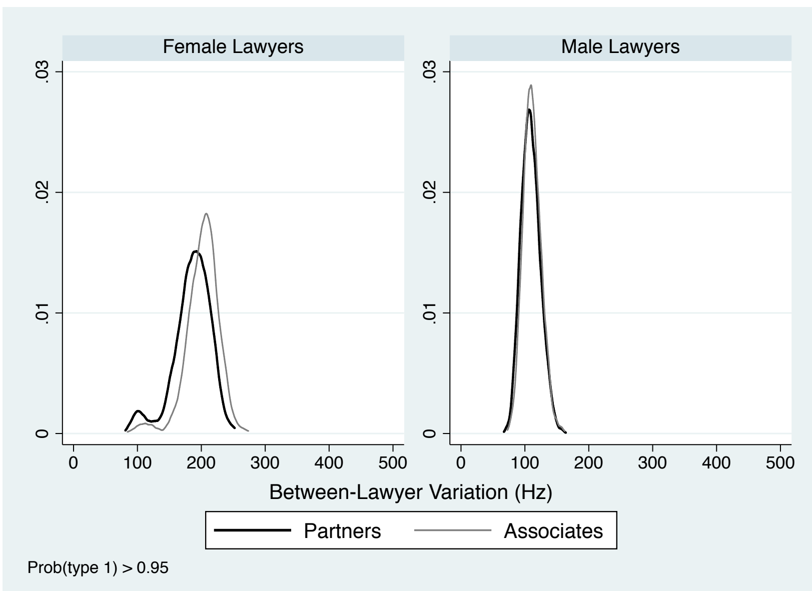

Figure 11: Between-Lawyer Densities by Title

Notes: This figure presents between-lawyer kernel density estimates of mean voice frequencies by gender and title using type 1 voicemail greetings that meet the high-threshold classification. The data used to estimate the densities are the mean voice frequency per voicemail greeting, which is obtained from the first three seconds of the greeting.

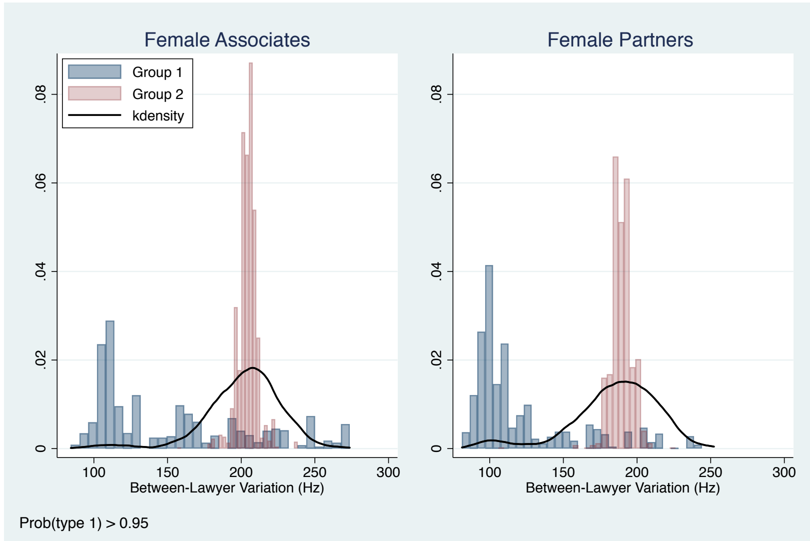

Figure 12: Between-Lawyer Unobserved Heterogeneity among Females

Notes: This figure presents histograms for each latent group of female lawyers based on predicted values from a multinomial logit model. Regressions are mean voice frequency of type 1 female lawyers, obtained from the first three seconds of the voicemail greeting, on firm fixed effects by title (associate or partner). The data used are the sample of type 1 voicemail greetings that meet the high-threshold classification (i.e., Prob(type 1) > 0.95). The solid lines are kernel density estimates of female associates (lefthand-side) and female partners (righthand-side), respectively.

Table 1: Descriptive Statistics of Sample of Law Firms

Variable	Obs	Mean	Std. Dev.	Min	Max
Firm rank	86	52.52	28.58	3	100
Profit per partner (million)	79	1.53	0.92	0.56	4.56
Revenue per lawyer (million)	79	1.01	0.25	0.55	1.68
Share female	83	0.36	0.03	0.25	0.44
Share minority	78	0.19	0.05	0.09	0.34
Share partners female	83	0.21	0.03	0.13	0.3
Share equity partners female	73	0.17	0.03	0.1	0.25
Total lawyers	81	972.84	779.73	82	4607
Associate-partner ratio	85	1.65	0.95	0.58	4.37
Number of offices	86	19.57	19.46	1	137
Lawyers per office	81	65.66	41.41	21.33	331
Profit margin	80	0.38	0.1	0.13	0.7
Revenue rank	67	45.88	27.97	1	98
Total revenue (billion)	81	0.93	0.62	0.18	2.65
Year established	86	1920.31	50.79	1792	2014

Notes: These data are on the final 86 Vault 100 firms in the sample collected from a number of external sources, including Vault.com. Productivity measures come from the Global 100 2016 published by Legal Business, including total gross revenue, which is used as an alternative method for ranking law firms. Data on lawyer counts and gender composition in 2016 come from the Law360 400 and the ATL Law Firm Gender Diversity Database, respectively. Since not all firms disclose these data, some of the variables in the table do not add up to 86.

Table 2: Voice Frequency Data by Lawyer Gender and Title

	Associate		Counsel/Other		Partner		All	
Gender	Freq.	Percent	Freq.	Percent	Freq.	Percent	Freq.	Percent
Female	7,442	44.71	2,413	39.02	3,904	22.78	13,759	34.43
Male	9,204	55.29	3,771	60.98	13,233	77.22	26,208	65.57
Total	16,646		6,184		17,137		39,967	

Notes: This table presents the number of voicemail greetings in the final sample of recordings I use in the analysis from each lawyer gender-by-title group.

Table 3: Machine Learning Voicemail Greeting Classification by Gender and Threshold

	Prob(type 1) > 0.95		Prob(type 0) > 0.95		Prob(type 1) > 0.50		All	
Gender	Freq.	Percent	Freq.	Percent	Freq.	Percent	Freq.	Percent
Female	3,715	25.86	3,551	28.17	7,550	35.27	13,759	34.43
Male	10,651	74.14	9,065	71.83	13,855	64.73	26,208	65.57
Total	14,366		12,616		21,405		39,967	

Notes: This table summarizes the distribution of voicemail greetings by gender and the probability that the lawyer recorded the greeting (type 1) or not (type 0). The probability estimates are based on results from a machine learning classification model (XGBoost). Details on the classification method are available in the text and in more detail in Materials and Methods.

Table 4: Between-Lawyer Unobserved Heterogeneity among Females by Title
Multinomial Logit Regression Results with Two Latent Groups

	Associates		Partners	
	Share	Mean Voice Frequency	Share	Mean Voice Frequency
Group 1	[0.050 0.078]	[153.17 153.36]	[0.063 0.089]	[121.39 121.41]
Group 2	[0.921 0.949]	[204.10 204.53]	[0.910 0.936]	[189.22 189.37]
Total	1,765		1,291	

Notes: This table presents results of estimating a multinomial logit model with two latent groups. Regressions are mean voice frequency, obtained from the first three seconds of the voicemail greeting, on firm fixed effects by title (associate or partner). The data used are the sample of type 1 voicemail greetings that meet the high classification threshold (i.e., Prob(type 1) > 0.95). Numbers in brackets represent 95 percent confidence intervals. Delta method standard errors are adjusted for clustering at the firm level. The results in this table suggest that about 5 to 9 percent of female lawyers are identified by the model as coming from a different distribution of mean voice frequency than the other female lawyers. Likewise, a slightly higher share of Group 1 are partners relative to Group 2.

A Supplementary Materials

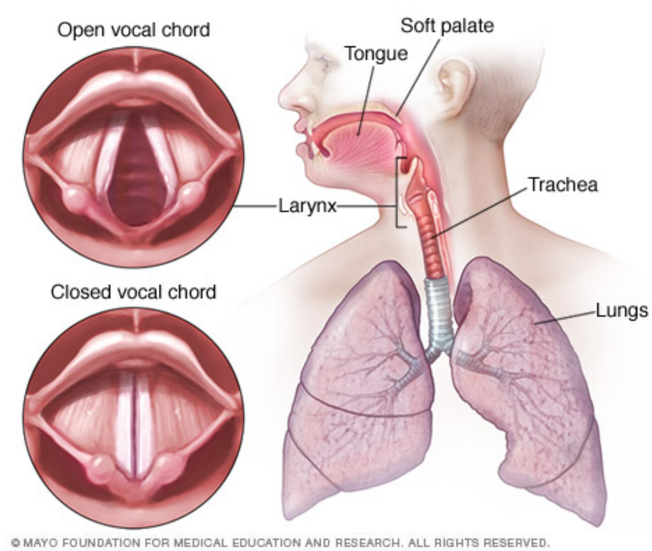

Figure A1: Voice Frequency and Vocal Anatomy

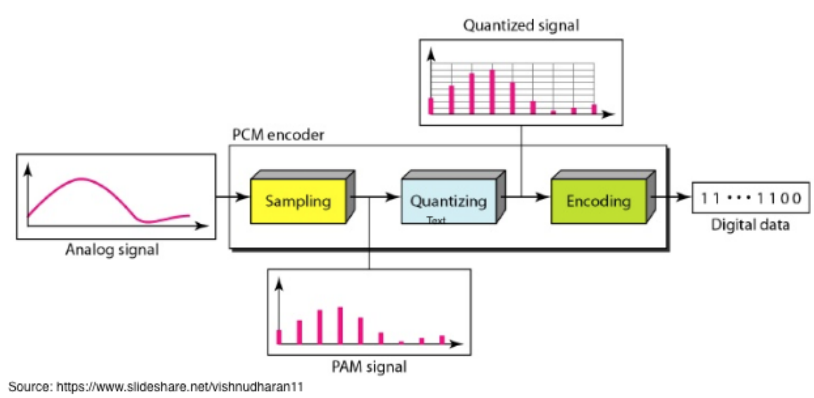

Figure A2: Illustration of Analog to Digital Audio Data Transformation

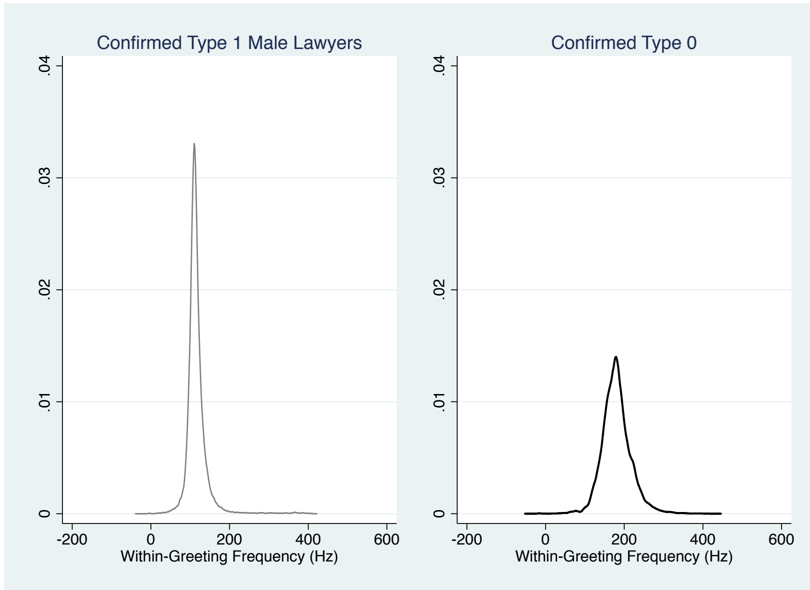

Figure A3: Primary Voice Frequency Modes of Males and Females

Notes: This figure presents kernel density estimates from voice frequencies from a pilot sample. The data used to estimate the densities are 100 demeaned voice frequency quantiles per voicemail greeting. The lefthand-side figure uses voicemail greetings recorded by male lawyers (n = 813), whereas the righthand-side figure uses voicemail greetings recorded in third person by females (n = 388).

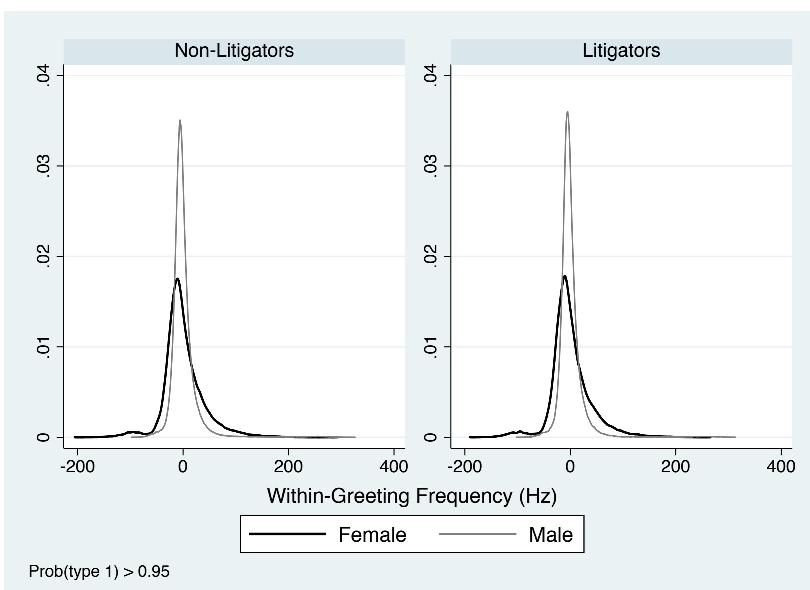

Figure A4: Lawyer Densities by Practice Area

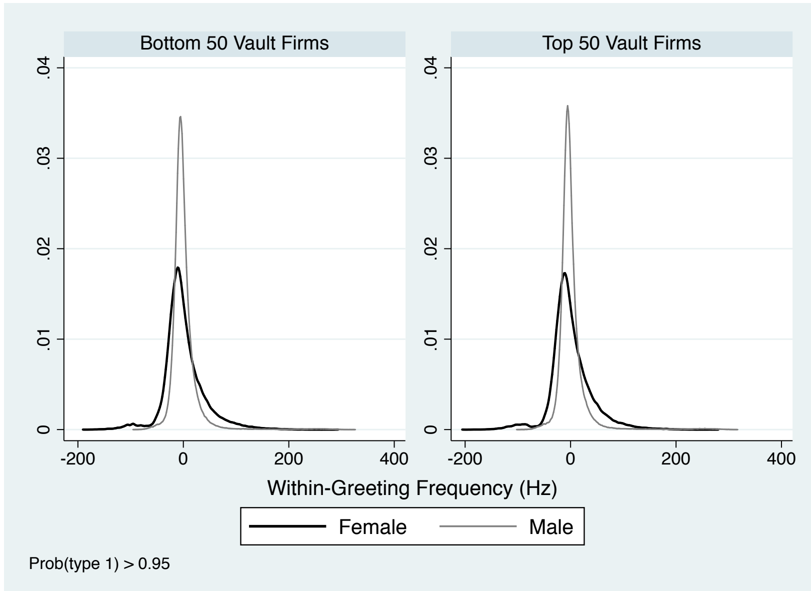

Figure A5: Lawyer Densities by Vault Firm Rank

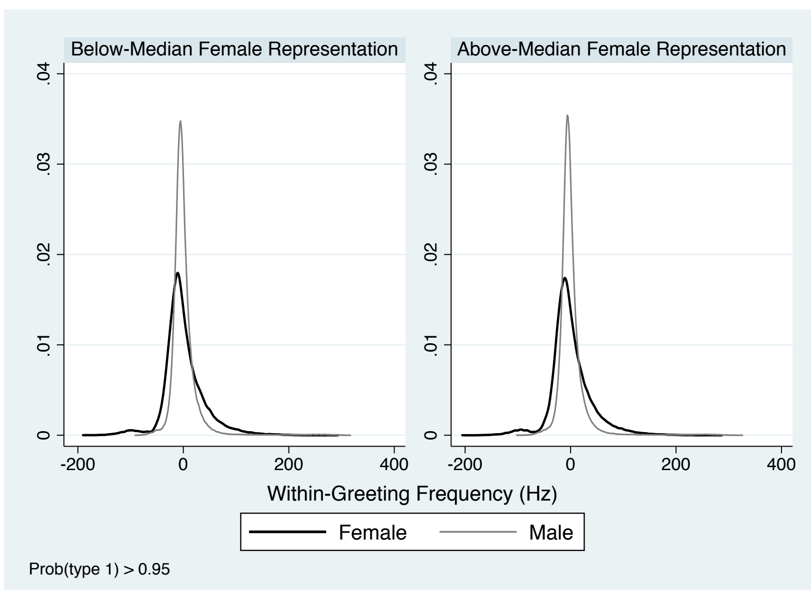

Figure A6: Lawyer Densities by Share Partners Female

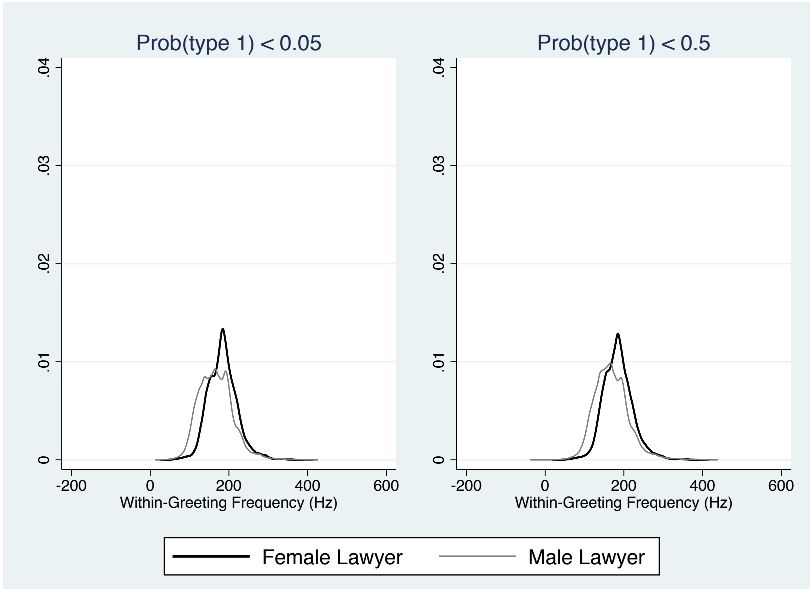

Figure A7: Lawyer Densities of Predicted Type 0 Greetings by Gender

Notes: This figure presents kernel density estimates of voice frequencies from type 0 voicemail greetings of male and female lawyers (i.e., recorded in third person by females). The data I use to estimate the densities are 100 greeting-demeaned voice frequency quantiles per voicemail greeting, which are obtained from the first three seconds of each greeting. The lefthand-side figure includes 12,616 type 0 voicemail greetings that meet the high-threshold classification (i.e., Prob(type 0) > 0.95), and the righthand-side figure includes 18,562 type 0 voicemail greetings that meet the low-threshold classification (i.e., Prob(type 0) > 0.5). Voicemail greetings of female lawyers constitute approximately one third of each sample.

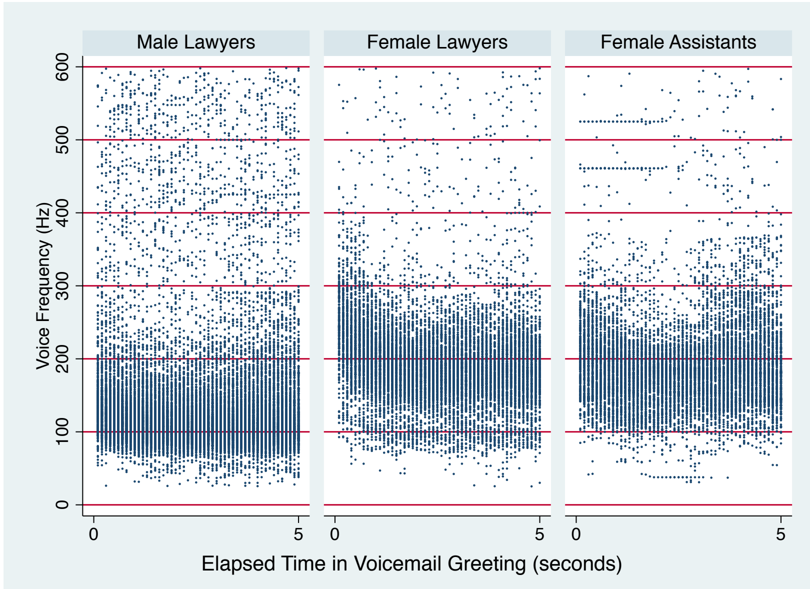

Figure A8: Voice Frequency over Time

Notes: This figure presents voice frequencies from a pilot sample of voicemail greetings recorded by male lawyers (*n* = 813) and female lawyers (*n* = 364) and female assistants (*n* = 388). The data used are 50 voice frequency samples selected from the first five seconds of each greeting at fixed intervals of 0.1 seconds. The floor was set to 25 Hz and the ceiling to 600 Hz. The mean voice frequency of a lawyer in each interval is represented by a dot in the scatter plot.

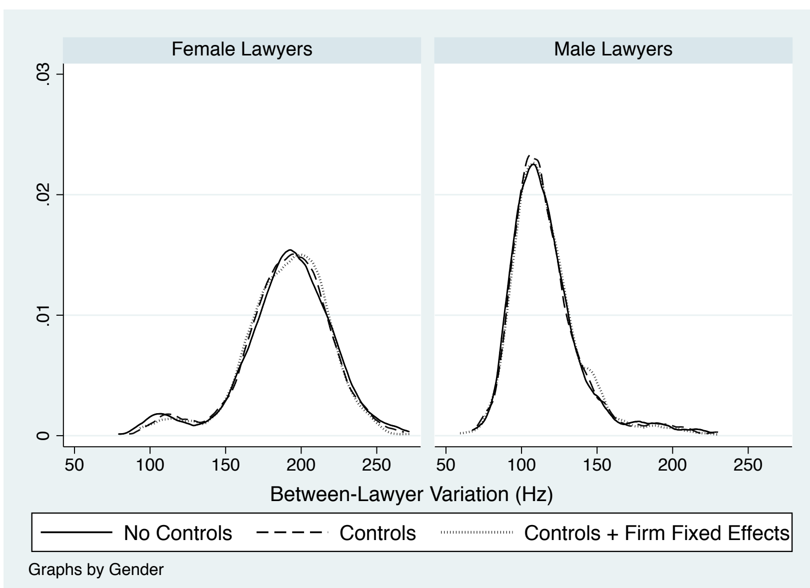

Figure A9: Sensitivity of Between-Lawyer Densities to Covariates

Notes: This figure plots kernel densities of type 1 voicemail greetings from the pilot sample. Dashed and dotted line kernel density estimates are based on residuals from regressions of a lawyer's mean voice frequency on lawyer gender, title, practice area, firm rank, law school rank, undergraduate school rank, age and years of experience. See Appendix B for details.

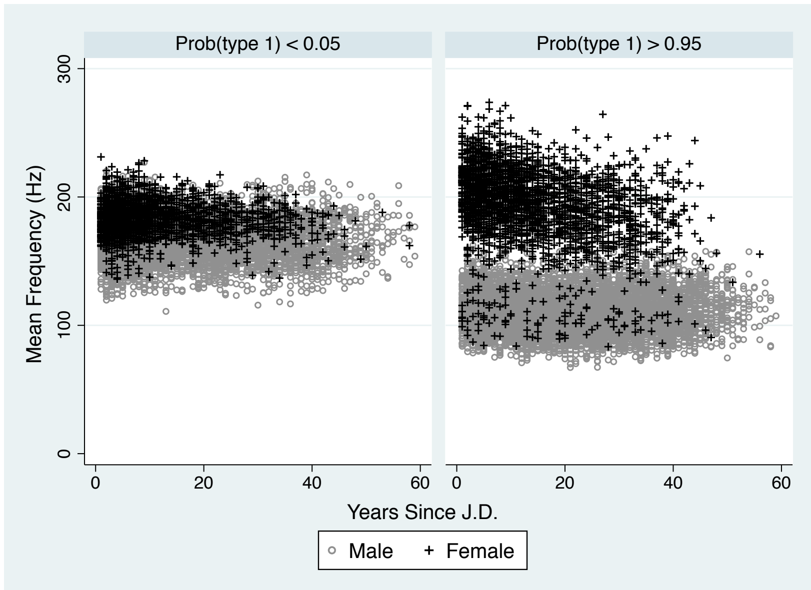

Figure A10: Between-Lawyer Scatter Plots by Years of Experience

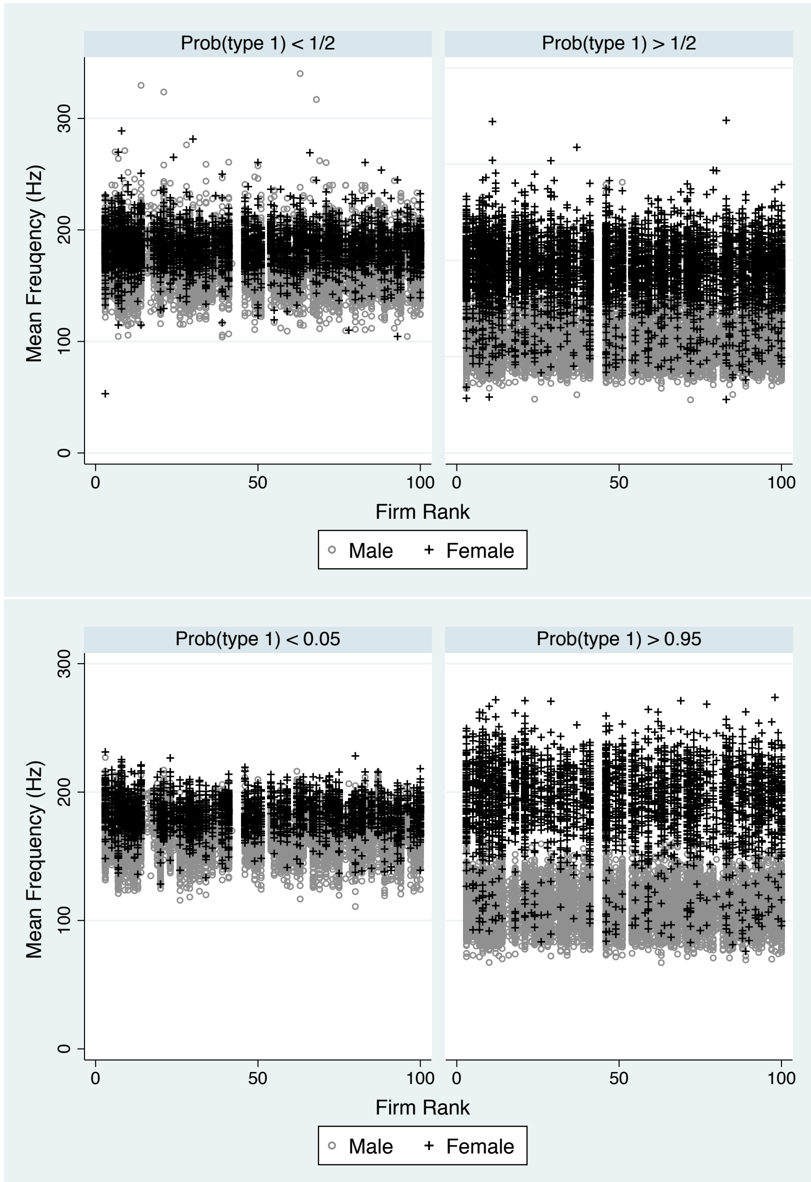

Figure A11: Between-Lawyer Scatter Plots by Predicted Greeting Type

Table A1: Three-Way Classification of Voicemail Greetings by Type and Gender

	Female Lawyers		Male Lawyers		All Lawyers	
Greeting Type	Freq.	Percent	Freq.	Percent	Freq.	Percent
Third person	96	14.41	292	18.38	388	17.21
Lawyer	364	54.65	813	51.16	1,177	52.20
Mixed	206	30.93	484	30.46	690	30.60
Total	666		1,589		2,255	

Table A2: Verbal Content of Voicemail Greetings
Most Common Words

	Type 0	Type 1	t statistic
reached	0.04	0.24	-14.03
take	0.01	0.08	-8.12
please	0.07	0.21	-9.96
leave	0.10	0.20	-6.26
message	0.13	0.14	-0.17
hello	0.02	0.11	-9.08
hi	0.01	0.19	-15.52
available	0.38	0.07	18.92
sorry	0.25	0.07	12.03
now	0.01	0.08	-8.01

Notes: This table contains verbal data statistics on a pilot sample of voicemail greetings that were successfully converted by Watson from 1,177 and 1,078 manually classified type 1 and type 0 greetings. The table lists the most common words used by either type 1 or type 0 lawyers in the first three seconds of their voicemail greeting. The numbers in each type column represent the share of greetings that contain each of the words. The statistics are based on IBM's Watson speech to text conversion application. The average number of words successfully recovered in type 1 greetings during the first three seconds is 5.41, whereas the number for type 0 recordings is 3.40 words (t-stat = 33.51). The list contains words identified by the application in at least 5 percent of type 1 or type 0 greetings. The complete least of words and details of the conversion exercise on the complete dataset are available in Materials and Methods.

B Description of Pilot Sample

I briefly describe part of the sample collected to pilot the methodology before applying it to the full sample.

For each of the Vault 100 firms in 2017, the procedure was to first collect the phone numbers of the first 10 alphabetically-listed litigation partners from each firm's webpage. I next used a VoIP-based communications software program developed by Voicent to dial the office numbers of these partners and record the first few seconds of each call. To maximize the chances of reaching the voicemail, I ran the code over several late weekend nights, and then classified the recordings by type. I found in 86 firms at least one voicemail greeting was personally recorded by a lawyer. These 86 firms comprise the baseline sample of firms. In

I next attempted to collect another 30 phone numbers from these 86 firms: 10 litigation associates, 10 tax partners and 10 tax associates. I chose these lawyers alphabetically using the same procedure as for the litigation partners. In total, I obtained 2,327 phone numbers from 86 firms. Of these, I obtained 2,255 recording, and focus my analysis on the subset of 1,177 self-recorded voicemail greetings of lawyers (type 1). There are 364 female and 813 male such lawyers in this sample. As seen in Table A1, almost half the recordings do not fall in this category. Some firms have voicemail greetings prepared by designated firm personnel or phone service operators (type 0). This class of recordings account for about 15 percent of the recordings. Other greetings include both the lawyer, typically only to personally identify by name, and a non-lawyer third person female voice (type 2).

Demographic data on type 1 lawyers come from their webpages on their firm's site. In most cases, there is information on law school, undergraduate school, including year of graduation, which I use to estimate age and experience. School names were matched with their *U.S. News & World Report* education rankings in 2017. Most lawyers are concentrated in coastal states, such as New York and California, plus Washington DC.

B.1 Voicemail Greeting Type and Lawyer Covariates

To examine whether lawyer demographics predict the type of voicemail greeting, I run regressions where the dependent variable is an indicator equal to 1 for a personal greeting and equal to 0 otherwise, and $\mathbf{x}$ is a vector of lawyer covariates. The results are in Table B1. The results suggest that male lawyers are 5 to 7 percent less likely to create a personalized voicemail greeting than female lawyers. Likewise, partners are at least 10 percent more likely to record a personal greeting than associates. There is some evidence for litigators being more likely to leave a personal greeting than tax attorneys; however, this pattern does not hold in general. And a similar argument can be made for the role of age and firm rank in the type of greeting. To sum, the data consistently show, even within firms, significant differences in greeting type by gender and title of lawyers.

B.2 Mean Voice Frequency and Lawyer Covariates

For a more detailed examination of different moments of the audio data, I present summary statistics of type 1 lawyers in Table B2. The average median frequency is about 5 Hz lower than the average means, where the median is 112 Hz and 184 Hz for male and female lawyers, respectively.

To examine the relationship between lawyer characteristics and the mean voice frequency, I run regressions where the dependent variable is the mean voice frequency (in Hz) of lawyer i in the first 3 seconds of the voicemail greeting. Standard errors are clustered at the firm-level. Table B3 presents results from regressing the mean voice frequency of type 1 lawyers on lawyer covariates. I find differences by gender and title, where columns 1-4 indicate about a 70 Hz lower mean voice frequency for males relative to females, and 10 Hz lower mean voice frequency for partners relative to associates. These results are robust to the inclusion of controls and firm fixed effect. There is some evidence for differences by practice group, where the mean voice frequency of tax lawyers is 3 Hz higher than of litigators. Columns 5 and 6 present results separately for each gender, where the estimates suggest that differences by title and practice group are primarily driven by female lawyers. The estimated differences among males are smaller and statistically insignificant. No other covariates predict the mean voice frequency with the exception of age, where the estimate of about 0.03 Hz lower mean voice frequency per year is economically insignificant. Relatedly, observables appear to explain far more variation in mean voice frequency of female lawyers (50 percent) than of male lawyers (20 percent), as indicated by the R^2 in columns 5 and 6. Given these results, I continue to focus on heterogeneity in gender and title.

B.3 Verbal Content of Voicemail Greetings

To get a sense of whether verbal content on the voicemail greetings is differentiated by gender, I compare the choice of words by type 1 lawyers. To do this, I apply IBM's Watson speech to text API to the first 3 seconds of each voicemail greeting. Excluding names, males say 0.3 more words than females (5.5 vs 5.19) during this 3-second period. Further, among the set of words used in at least 5 percent of the clips, male lawyers are statistically significantly more likely to say "sorry" (8 vs 4 percent), "please" (24 vs 15 percent), "leave" (22 vs 13 percent), "message" (16 vs 9 percent), and "take" (9 vs 5 percent), whereas female lawyers have a higher frequency of saying "hello" (15 vs 9 percent). Overall, I do not find meaningful differences in verbal content across lawyers by gender.

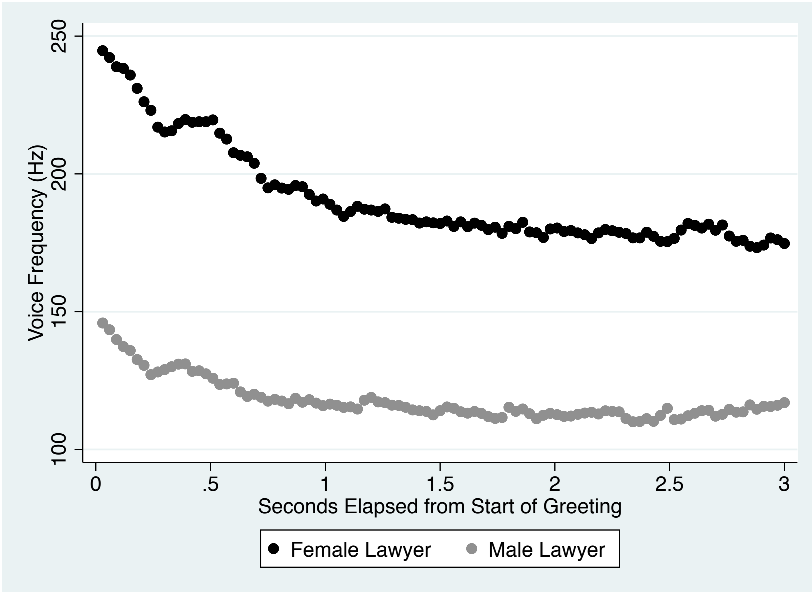

Figure B1: Average Voice Frequency over Time

Notes: This figure presents voice frequencies from a pilot sample of voicemail greetings recorded by male lawyers (n = 813) and female lawyers (n = 364). The data used are 100 voice frequency samples selected from the first three seconds of each greeting at fixed intervals of 0.03 seconds. In each interval, the mean voice frequency is averaged across all lawyers and is represented by a dot in the scatter plot.

Table B1: Voicemail Greeting Type and Lawyer Characteristics

	Dependent Variable: Type 1 Voicemail Greeting = 1; Otherwise = 0					
	(1)	(2)	(3)	(4)	(5)	(6)
Male	-0.0731***	-0.0670***	-0.0502**	-0.0643***	-0.0714***	-0.0623***
	(0.0221)	(0.0218)	(0.0232)	(0.0230)	(0.0252)	(0.0225)
Associate	-0.143***	-0.173***	-0.187***	-0.0905***	-0.107**	-0.149***
	(0.0283)	(0.0264)	(0.0309)	(0.0261)	(0.0428)	(0.0423)
Practices tax	-0.0162	-0.0570***	-0.0570***	-0.0366	0.0152	-0.0339
	(0.0235)	(0.0207)	(0.0214)	(0.0233)	(0.0278)	(0.0246)
Firm rank					0.00215***	
					(0.000645)	
Law school rank					-0.000359	-0.000405
					(0.000366)	(0.000355)
College rank					0.000201	-0.000141
					(0.000314)	(0.000311)
Age					0.000425***	0.0000721
					(0.000101)	(0.0000980)
Years experience					-0.00259	-0.00273
					(0.00164)	(0.00165)
Constant	0.638***	0.662***	0.764***	0.844***	0.779***	0.947***
	(0.0268)	(0.0219)	(0.0254)	(0.0210)	(0.0692)	(0.0364)
Firm fixed effects		X	X	X		X
Excluded greetings			Machine	Mixed	Mixed	Mixed
Observations	2255	2255	1867	1565	1387	1387
R^2	0.020	0.165	0.186	0.216	0.041	0.243
Firms	92	92	88	92	91	91

Notes: This table presents OLS regression results. The dependent variable is equal to 1 if the voicemail greetings is type 1, and is equal to 0 otherwise (type 0 or type 2) based on the three-way manually classified method described in Appendix B. Standard errors adjusted for clustering at the firm-level in parentheses.* p < 0.10, ** p < 0.05, *** p < 0.01.

Table B2: Selected Voice Frequency Moments of Type 1 Voicemail Greetings

	Male Lawyers		Female Lawyers	
	Mean	Std. Dev.	Mean	Std. Dev.
Mean	116.34	23.74	189.71	32.46
Min	75.87	22.55	107.61	43.56
Q1	103.68	20.92	169.30	30.50
Median	111.55	23.15	184.31	33.06
Q3	123.84	29.50	207.66	38.12
Max	214.29	90.12	294.02	58.30
Observations	813		364	

Notes: This table presents summary statistics of the voice frequency obtained from a pilot sample of type 1 manually-classified voicemail greetings.

Table B3: Mean Voice Frequency of Type 1 Voicemail Greetings and Lawyer Characteristics

Dependent Variable: Mean Voice Frequency (Hz)						
					Male	Female
	(1)	(2)	(3)	(4)	(5)	(6)
Male	-71.13***	-70.65***	-72.34***	-72.02***		
	(2.244)	(2.431)	(2.423)	(2.595)		
Associate	10.07***	11.19***	11.75***	13.92***	4.607	16.78***
	(1.425)	(1.635)	(2.433)	(2.812)	(3.184)	(5.823)
Practices tax	3.071**	3.647**	3.116*	4.040**	2.059	10.02**
	(1.487)	(1.676)	(1.674)	(1.906)	(1.907)	(4.117)
Firm rank			0.0396			
			(0.0267)			
Law school rank			0.0266	0.0242	-0.0241	0.0708
			(0.0237)	(0.0268)	(0.0293)	(0.0583)
College rank			0.00224	0.000635	0.0386	-0.00708
			(0.0214)	(0.0231)	(0.0233)	(0.0427)
Age			-0.0307***	-0.0329***	-0.0398***	0.644
			(0.00581)	(0.00643)	(0.00565)	(0.986)
Years experience			0.160	0.183	0.144	-0.859
			(0.106)	(0.120)	(0.113)	(1.030)
Constant	183.5***	182.6***	178.8***	179.4***	110.2***	165.4***
	(2.005)	(1.986)	(3.828)	(3.991)	(3.325)	(27.21)
Firm fixed effects		X		X	X	X
Observations	1177	1177	1054	1054	722	332
R^2	0.630	0.656	0.645	0.676	0.195	0.505
Firms	86	86	84	84	84	73

Notes: This table presents OLS regression results. The dependent variable is the mean voice frequency based on the first three seconds of type 1 lawyers' voicemail greetings. Standard errors adjusted for clustering at the firm-level in parentheses.* p < 0.10, ** p < 0.05, *** p < 0.01.

Table B4: Verbal Content of Type 1 Voicemail Greetings by Gender

	Male Lawyers	Female Lawyers	t-stat
please	0.24	0.15	3.77
reached	0.23	0.25	-1.03
leave	0.22	0.13	3.84
hi	0.19	0.19	0.12
message	0.16	0.09	3.13
take	0.09	0.05	2.18
hello	0.09	0.15	-2.61
now	0.09	0.07	1.07
sorry	0.08	0.04	2.68
available	0.07	0.08	-0.93
office	0.04	0.06	-1.46
voicemail	0.04	0.05	-0.75

Notes: The table lists the most common words used by either male or female lawyers in the first three seconds of their voicemail greeting. The numbers in each gender group column represent the share of greetings that contain each of the words. The statistics are based on IBM's Watson speech to text conversion application. This table contains verbal data statistics on a sample of 839 voicemail greetings that were successfully converted by Watson from 1,177 manually classified type 1 greetings from a pilot sample. The average number of words recorded by male lawyers in the first three seconds is 5.5, whereas for female lawyers is 5.19 words (t-stat = 3.27). The list contains words identified by the application in at least 5 percent of the female or male greetings. The complete least of words and details on the conversion exercise on the complete dataset are available in Materials and Methods.

C Quality of Voice Frequency Data

C.1 Background

The first thing to consider in measuring and analyzing voice is its inherent transformation from continuous analog data to discrete digital data. Pulse-code modulation (PCM) is the most common method used to digitize analog signals. Thus, the first layer of noise introduced in the process of data collection is the discretization error associated with the transformation of sound quality that is continuous (e.g., face-to-face conversations, live music) to a series of ones and zeros that depend on the sampling rate–the number of observations obtained per unit of time. In addition, the finite resolution of digitized voice introduces a quantization error, which is proportional to the bit depth–the dimensionality of each observation. Both minimize loss of information; however, the process introduces measurement error, where secular features of the original sound distribution may be lost. Figure A2 illustrates the PCM process from continuous to discrete audio signal. The greater the dimensionality of the bit, the more detailed information describing the signal range of sound can be stored. The discretization error stems from discretizing a continuous time domain, whereas the quantization error stems from discretizing a continuous amplitude range.[30]

The second layer of measurement error occurs when data packets travel over the Internet back to the caller. Recordings for the training sample were obtained using a single VoIP channel whereas for the main sample I utilized 50 VoIP channels to simultaneously calls blocks of 50 numbers at a time. Since the network is shared with other users, there is protocol on packet loss. VoIP uses a Realtime Transport Protocol (RTP) that guides the packets to their destination, but prioritizes timely delivery of the packets to prevent loss of sound quality. These packets are sent every 20 to 30 milliseconds containing the digitized audio samples. For example, given a standard sampling rate of 8000 pulses per second, one second of digitized audio travels in 50 packets of 160 samples. The dimensionality of the data (i.e., bit depth) determines the load on each data packet and influences its ability to maneuver through congestion on the Internet. If a packet is dropped because the connection lags or is temporarily lost, then the audio information cannot be resent. Typically, packets experiencing the worst delays will be dropped resulting in low latency at the cost of data loss. In sum, some fraction of data packets containing the digitized voicemail greetings is dropped before reaching its destination. In particular, in congested networks, increasing sample size may be undermined by lower data quality.

C.2 Quality Assessment

Recordings for the pilot sample were obtained using a single VoIP channel whereas for the main sample I utilized 50 VoIP channels to simultaneously calls blocks of 50 numbers at a time. Because sound quality is a major issue in the telephony industry around the world, several measures are recognized internationally to assess service quality.[31] To asses the sound quality of the audio recordings, I used a fully-referenced method known as "P.862" on a set of 855 voicemail greetings that were recorded twice, once in the pilot sample and once in the full sample. To assess the quality of the full sample recordings, the pilot sample recordings were designated as the reference signal.

To ensure that the data quality analysis extends beyond the "overlapping" sample to the entire sample I use, I show in Figure C1 kernel density estimates of the mean voice frequencies (i.e.,

[30]The discretization error is the difference between $f'(x)$ and its approximation $\frac{f(x+h)-f(x)}{h}$, where h equals inverse of the sampling rate called the sampling period (e.g., for the standard 8K sampling rate, h = 1/8 milliseconds). The quantization error is the rounding error associated with difference between the actual values of a continuous function f and the step function that approximates f constructed from a finite discrete set of values determined by bit depth. In particular, the Signal to Quantization Noise Ratio (SQNR) goes up by about 6 dB for each bit added to the sample. This increase in the power of the waves allows for lower ranges of sounds to be captured and distinguished from noise.

[31]See more on the Telecommunication Standardization Sector here: https://www.itu.int/en/ITU-T/Pages/default.aspx

between-lawyer densities) for both the overlapping and non-overlapping samples separately. The figure confirms that there are no significant differences between these densities. [32]

The scores are on a relative scale and built upon many factors which can affect voice quality. For example, delay in data propagation, which is the time required for a digital signal to travel from origin to destination across the entire network, is a key difference between the pilot and full up data. See Figure C2 and Table C1 for results from this analysis. In particular, the quality estimate is highly correlated with the absolute difference in mean frequency estimates from both samples. Further, the full up observations that are estimated to be above median quality produce very similar estimates to the ones obtained with the pilot sample observation. However, those that are below median quality produce different and insignificant estimates, but the pilot sample estimates are still robust. This provides evidence on a connection between audio quality and the ability to detect secondary moments in the mean voice frequency distribution.

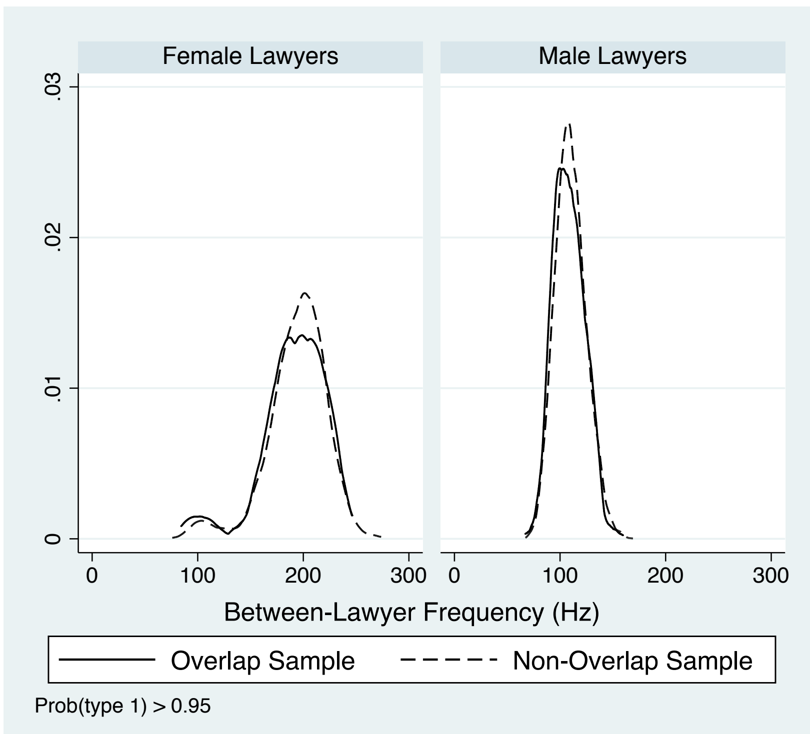

Figure C1: Comparison between Full Sample and Audio Quality Assessment Sample

[32]The estimates in Figure C1 are based on 440 and 13,925 observations in the overlapping and non-overlapping sample, respectively. One key issue to consider is that the overlapping sample comprises confirmed type 1 recordings, whereas the non-overlapping sample includes all recordings with $\Pr(\text{type } 1) > 0.95$. This means that the latter include some fraction of third-party recordings. This results in a slight rightward shift of the non-overlapping relative to the overlapping density for men, and a slightly less bimodal density of the non-overlapping relative to overlapping sample for females.

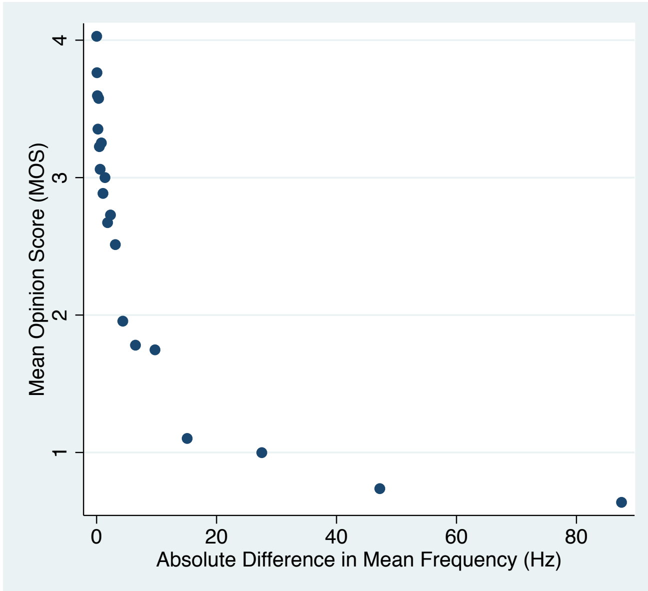

Figure C2: Audio Quality Measure and Differences in Acoustics between Recordings

Table C1: Audio Quality Comparison of Voicemail Greeting Recordings by VoIP Application

Dependent Variable: Mean Frequency of Voicemail (Hz)

	ML	Full	ML	Full	ML	Full
	(1)	(2)	(3)	(4)	(5)	(6)
	All Clips		Above Median Quality		Below Median Quality	
Male	-72.79***	-73.57***	-79.35***	-78.54***	-65.67***	-68.53***
	(2.484)	(2.013)	(2.569)	(2.638)	(3.967)	(2.828)
Associate	10.10***	5.849***	9.065***	8.928***	12.15***	4.894
	(1.613)	(1.850)	(2.165)	(2.144)	(2.647)	(3.168)
Practices tax	3.550**	0.151	3.617	4.026*	4.075	-1.395
	(1.736)	(2.094)	(2.202)	(2.190)	(2.571)	(3.155)
Constant	184.1***	191.2***	188.6***	187.8***	178.8***	192.3***
	(2.194)	(1.878)	(2.451)	(2.521)	(3.588)	(2.879)
Observations	855	855	429	429	426	426
R^2	0.664	0.640	0.762	0.756	0.572	0.547
Firms	75	75	62	62	74	74
F	327.037	556.806	397.927	370.121	126.575	207.998

Notes: This table compares a subset of voicemail greetings that were recorded twice: once over a single VoIP channel and another time as part of the 50 simultaneous VoIP channels I used to collect the full sample. The audio quality of each recording is based on the fully-references method described in B.2. Standard errors adjusted for clustering at the firm-level in parentheses.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

D Using Acoustic Data to Predict Lawyer Observables

This paper has focused on voice frequency. However, alternative measurements and representations of voice patterns, including the intensity of voice and the Fourier transform of audio data can be useful in testing whether a lawyer's voice predicts lawyer characteristics .[33] This approach is the opposite of using lawyers characteristics to predict voice. For example, to the extent that voice data can effectively discern among different types of workers, one might wonder whether clustering lawyers solely based on their voices can uncover key differences among workers.

For this exercise, I used an unsupervised machine learning algorithm. This method is similar in spirit to principal component analysis. In a two-step process, machine learning is used to (1) find the optimal number of clusters and (2) estimate the location (i.e., coordinates) of each lawyer with reference to the clusters' centroids. The results suggest that two clusters best explain the data (see Figure D1). The classification of lawyers into two groups varies by the type of audio data given to the machine. To summarize the results, Table D1 presents a mapping between the clusters and a set of (dichotomized) lawyer characteristics, where each number represents the percent of lawyers classified by these clusters into the appropriate demographic group. Interestingly, the correlation between the clusters produced by unsupervised machine learning and lawyer gender is high but not perfect, suggesting the machine's inability to identify other groups in the population (specifically female lawyers who utilize the "male" mode). Because the clustering method involves minimizing a distance function between observations and a cluster centroid, the mean and variance of the data may receive excessive weight in clustering observations and ignore secondary moments of the data in the prediction process. I provide further details and results of these analyses, including separate clustering exercises for male and female lawyers, in Materials and Methods.

[33] In a nutshell, a Fourier analysis transforms audio data from a time to frequency domain using a sequence sinusoidal functions. Since voice data are more complex than sinusoidal sequences the transformation of discretized audio data introduces discontinuities when using a finite number of points to calculate the Discrete Fourier Transform (DFT) using a Fast Fourier Transform (FFT) algorithm. The FFT typically computes the DFT precisely up to a floating-point error.

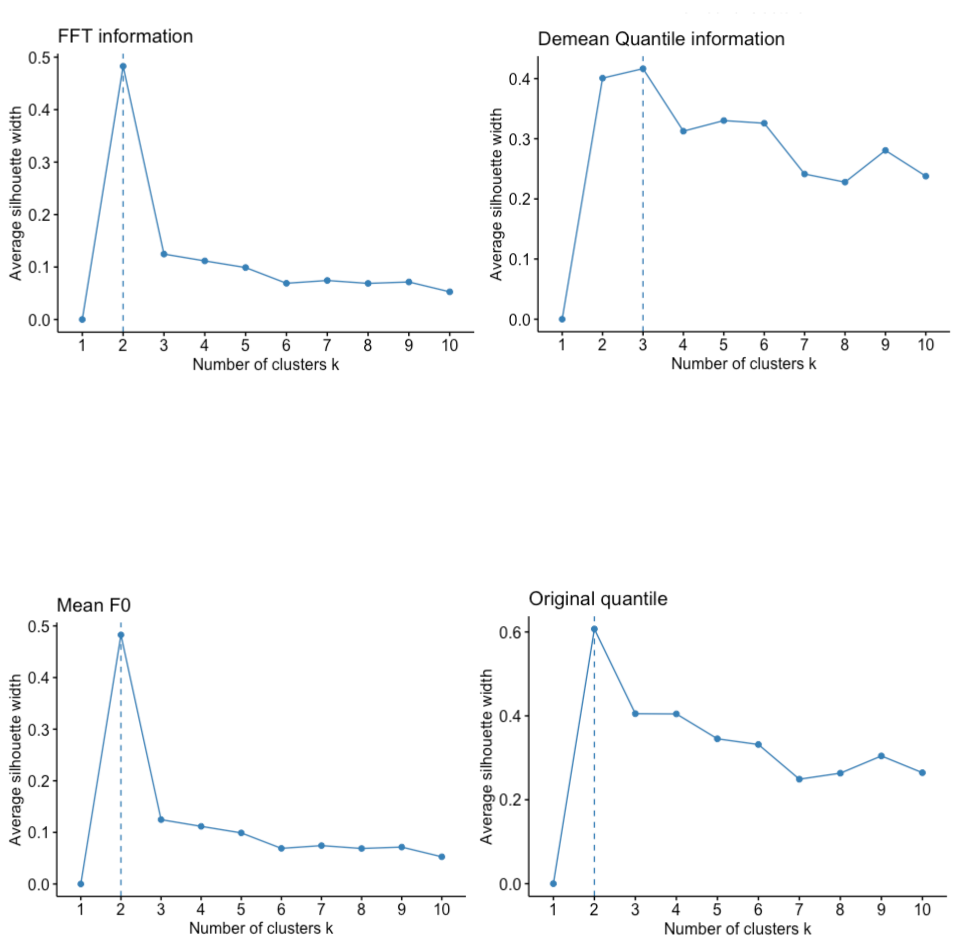

Figure D1: Optimal Number of Clusters by Type of Acoustic Data

Table D1: Unsupervised Clustering of Lawyers by Type of Acoustic Data

Percent of Overlap Between Estimated Clusters and Lawyer Covariates

Acoustic Data:	Quantiles	Mean Frequency	Demeaned Quantiles	FFT
Male	92%	92%	71%	60%
Associate	65%	65%	59%	57%
Practices tax	55%	55%	55%	56%
Firm rank	53%	52%	50%	55%
Law school rank	51%	51%	53%	51%
Age	61%	61%	54%	54%
Years experience	62%	62%	55%	54%

Notes: This table presents correlations between lawyer covariates and unsupervised two-way clustering results based on acoustic data obtained from the first three seconds of each voicemail greeting. The data are 1,177 confirmed type 1 voicemail greetings from the pilot sample. Each cell represents the maximum percent of lawyers correctly classified by the acoustics-generated clusters into one of two groups within each lawyer covariate . All Covariates are binary, where continuous covariates were dichotomized into above- or below-median value in the sample. For example, the top left cell suggests that the machine's clustering classification based solely on 100 voice frequency quantiles per lawyer mimics gender classification at the rate of 0.92. Likewise, the machines classification based on Fast Fourier Transform data obtained from the voicemail greetings mimics gender classification at a rate of 0.6.