Online Appendix for

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

Yosh Halberstam (University of Toronto)

Correspondence to: yosh.halberstam@utoronto.ca

This PDF file includes:

Materials and Methods
Supplementary Text
Figures S1 to S9
Tables S1 to S11

Contents

Materials and Methods

I provide further details on the data collection process for each sample of workers (Section 1), estimating frequencies from audio data (Section 2), the machine learning classification process for estimating the probability that a given clip is self-recorded by the lawyer (Section 3), and the survey design and implementation to test human detection and perception of bimodal clips (Sections 4 and 5).

Section 1. Data Collection Process for Worker Samples

Section 1.1. Main Dataset: Lawyers from Vault 100 Law Firms

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

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

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

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

Section 1.2. Verified Female Lawyers from the Main Dataset

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

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

Section 1.3. Auxiliary Dataset: Law Firm Assistants

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

1. The lawyers associated with these clips were classified as male by ALM.

2. The clips had a mean frequency above 150 Hz, well above the unitary male frequency mode of 100 Hz.

3. The clips had a probability of being self-recorded of at least 0.25, as determined by my machine learning classification. (see Section 3 for details on the machine learning classification process).

I classified a clip as assistant-recorded if the clip contained third-person pronouns and was not recorded by a human. I eliminated clips that were recorded by non-executive assistants (i.e., the assistants who spoke generally on behalf of the firm rather than for a specific lawyer), clips that were fully or partially machine-recorded, and poor-quality clips. Through this process, I found 237 clips that were fully recorded by female executive assistants on behalf of female lawyers, and 412 clips that were fully recorded by female executive assistants on behalf of male lawyers.

Section 1.4. Auxiliary Dataset: Lawyers Who Switched Firms

I define "switchers" as lawyers who had switched firms since the initial collection of their voicemails. As noted in Section 1.1, the period of initial collection was between May 2017 to January 2018. To find the switchers, about two years after the period of initial collection, I recruited MTurk workers to check the webpages of lawyers whose voicemails had a probability of being self-recorded of at least 0.5, as determined by my machine learning classification (see Section 3 for details on the machine learning classification process). The MTurk workers were selected on a first-come-first-served basis from a pool of US-based MTurk workers who had a HIT approval rate of 99% or more and who had over 10,000 approved HITs.

I determined whether a lawyer had switched firms by clicking on their original profile page URL. If the URL linked to the lawyer's profile page, I determined that the lawyer had not switched firms. If the URL did not link to the lawyer's profile page, for instance redirecting to a general directory or an error page, I determined that the lawyer had switched firms. See Figure S4 for the survey format.

I then recruited MTurk workers to find the switchers' new firms, profile page URLs, and personal office phone numbers. I provided MTurk workers with the switchers' names and old firms, and instructed respondents not to use sites other than Google, LinkedIn, and the lawyer's new firm website when providing information to avoid third-party websites of unknown reliability. See Figure S5 for details on the survey format and instructions.

As in Section 1.1, I used Voicent to collect voicemails from the new phone numbers. For each lawyer, I checked if both the old and new recordings were self-recorded by the same lawyer following the procedure in Section 1.2. In total, I found 627 lawyers with self-recorded voicemails at both their old and new firms. Of these 627 lawyers, 198 lawyers were female, as further verified using the procedure in Section 1.2.

Section 1.5. Auxiliary Dataset: Promoted Lawyers

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

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

1. Confirmed to have stayed at their original firm from the survey in Section 1.4.

2. Associates as of May 2017 – January 2018.

3. Had voicemails with a probability of being self-recorded of at least 0.5 (see Section 3 for details on the machine learning classification process).

I provided MTurk workers with the lawyers' website URLs, and instructed them to classify the lawyers' job title based on their personal webpage profiles. See Figure S6 for the survey format. This process elicited promotions data for 1,925 female lawyers whose clips were verified as self-recorded by a female lawyer using the procedure in Section 1.2, and who were Associates as of May 2017 – January 2018. Of these 1,925 lawyers, 196 were promoted to Partner and 137 were promoted to Counsel/Other within the same firm as of January 2020.

Section 1.6. Auxiliary Dataset: Supreme Court Lawyers

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

Section 1.7. Auxiliary Dataset: RE/MAX Real Estate Agents

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

As in Sections 1.1 and 1.2, I used Voicent to collect voicemails from these phone numbers, and manually classified these clips as self-recorded by the agent by listening to them. I successfully collected 539 self-recorded voicemails from the residential real estate agents, and 527 self-recorded voicemails from the commercial real estate agents.

I then recruited MTurk respondents to check the gender of agents who had self-recorded voicemails. They were selected on a first-come-first-served basis from a pool of US-based MTurk workers who had a HIT approval rate of 99% or more and who had over 10,000 approved HITs. I provided MTurk respondents with the agents' names and directory URLs, and instructed them to classify the agents' gender based on their profile pictures. See Figure S7 for the survey format. Of the 539 residential real estate agents, 337 were classified as female. Of the 527 commercial real estate agents, 159 were classified as female.

Section 2. Estimating Frequencies from Audio Data

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

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

Pitch Floor: 50 Hz;

Pitch Ceiling: 400 Hz;

Time Step: 5 milliseconds;

Window: Hanning;

Silence Threshold: 0.03;

Voicing Threshold: 0.45;

Octave Cost: 0.01;

Octave-Jump Cost: 0.35;

Voiced/Unvoiced Cost: 0.14.

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

- *Pitch Floor:* The minimum frequency that will be considered for estimation. A pitch floor of 50 Hz is well below the range of human voice frequencies produced by the natural voice register. The pitch floor also determines the length of the analysis window, which is 3/*Pitch Floor* seconds long. With a pitch floor of 50 Hz, the analysis window is 0.06 seconds long.

- *Pitch Ceiling*: The maximum frequency that will be considered for estimation. A pitch ceiling of 400 Hz is well above the range of human voice frequencies produced by the natural voice register.

- *Time Step*: The interval between frequency estimates. The points in the clip where the first and the last frequency estimates are taken depend on the length of the analysis window, which in turn depends on the pitch floor. For a clip represented as a time interval [0, 3] in seconds, a pitch floor of 50 Hz and time step of 5 milliseconds mean that Praat produces one frequency estimate per 5 milliseconds in the subinterval

[0.03, 2.97], for a total of 589 frequency estimates including the endpoints of the subinterval.

I used the default Praat settings for the following:

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

with Octave-Jump Cost, this is a global path finder parameter that affects estimates across rather than only within frames. I use the default 0.14 as the voiced/unvoiced cost.

With these parameters, each clip comprises 589 frames. For each frame, the best frequency candidates were determined using the function "Get value in frame". Through this process, I obtained 589 frequency estimates for each clip, though many of these estimates are voiceless (i.e. missing values), and the number of such missing values differs across clips. See Supplementary Text for robustness checks on the values of the parameters used.

Section 3. Machine Learning Classification Process

As in Section 1.2, I define a self-recorded clip as a clip which contains only a lawyer's voice, as opposed to an automated voice, an assistant's voice, or a combination of voices. Due to the large number of clips in the main dataset of 39,962 lawyers, I used machine learning to predict the probability that a clip is self-recorded for the main dataset. I used text information, acoustic information, and demographic information as predictors.

I obtained text information for each clip by transcribing the clips with IBM Watson Speech Recognition API. I decomposed the transcribed text into individual words, and selected the 50 most frequent words and the number of words per clip as predictors (Table S2). I further manually selected 12 frequently occurring phrase patterns as predictors (Table S3).

I obtained acoustic information for each clip using Praat and selected eight acoustic variables as predictors (Table S4). I also obtained the demographic information of each lawyer from their websites. This information included: job title, practice area, law school, undergraduate school, any graduate degrees, graduation year for each degree earned, academic honors earned, and gender. Gender was assessed by scrutinizing lawyers' first names, subjectively classifying photos, searching for gendered pronouns in the lawyers' biographical descriptions, and cross-referencing the recorded greeting. Generally, these gender indicators perfectly corroborated one another.

I set aside about 10% of the sample (the "ML sample") for training, testing, and validation. I manually classified all clips in the ML sample as self-recorded or otherwise following the procedure in Section 1.2. I randomly selected 80% of the ML sample for training and validation (the "training and validation sample") and the remaining 20% for testing ("the testing sample"). To address any possible overfitting issues, I used 5-fold cross validation to tune the model with the objective of minimizing logarithmic loss. I applied this procedure to four machine learning models: random forest, support vector machine, *k*-nearest neighbors, and XGBoost. Table S5 compares the predictive accuracy of the four models on the testing sample using a probability threshold of 0.5, and shows that XGBoost significantly outperformed the other three models with a predictive accuracy of 93.52%. With a probability threshold of 0.95, the XGBoost model had a predictive accuracy of 99.12%. I thus used XGBoost as my final machine learning model.

Figure S7 ranks the fifteen most important attributes in the XGBoost model. For each decision tree, the importance of each attribute is calculated as the amount by which 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. Acoustic measurements played the most important role. The most important non-acoustic variable was the number of words spoken by the lawyer in the greeting (coming in at sixth place). Gender was the most important–and indeed the only–demographic variable identified in the top-fifteen ranking (coming in at eighth place).

Section 4. Survey Design for Detection of Mode-Switching

To check if it was possible to consciously discern mode-switching behavior with the human ear alone, I surveyed 200 MTurk workers and asked them to listen to paired clips with similar mean frequency. Each pair of clips comprised a Group 1 (bimodal) clip and a Group 2 (unimodal) clip, and workers were asked to detect the bimodal clip. The clips were played in reverse to eliminate the influence of vocal content while preserving the acoustic characteristics. The survey instructions and format are shown in Figure S8.

The 250 paired clips I used for the survey are a subset of verified female Associates, one from each group (see Section 1.2), from the main dataset. The clips obey the following criteria:

1. Mean frequency 175-225 Hz.
2. Demeaned low mode location estimate between -120 and -90 Hz (Group 1) or -40 and -20 Hz (Group 2).

Going from low to high mean frequency, I sequentially paired each Group 1 clip with a Group 2 clip that had a mean frequency within 1 Hz of the Group 1 clip. To obtain the final sample, pairs were chosen to minimize the mean frequency difference: 125 pairs where the bimodal clip had higher mean frequency, and 125 pairs where the bimodal clip had lower mean frequency.

In MTurk terminology, a HIT is a copy of the same survey with different inputs. For example, HIT 1 may ask respondents to listen to clips 1-50, and HIT 2 may ask respondents to listen to clips 51-100, with both HITs having the same format and instructions. For each HIT question, the position of bimodal clip was randomized as either Recording 1 or Recording 2, and across questions for each HIT, the bimodal clip appeared 25 times as Recording 1 and 25 times as Recording 2. In addition, across HITs, each of the 250 pairs was used 40 times. These specifications were not made known to the workers.

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

I sorted respondents into quartiles based on their mean accuracy on this survey. In a follow-up survey, I asked the lowest and highest quartiles of female respondents and the lowest and highest quartiles of male respondents to complete a shorter version of the original survey with 3 sections of 10 questions each. The base compensation of the follow-up survey was the same as that of the original survey, and the bonus compensation was doubled. 46 out of 50 male respondents and 38 out of 50 female respondents completed the follow-up survey. As such, the response rate was 92% for males, 76% for females, and 84% in aggregate.

Section 5. Survey Design for Relative Rating of Unimodal and Bimodal Clips

To check if bimodality is correlated with perceptions of lawyers' characteristics, I surveyed 200 MTurk workers and asked them to listen to the same paired clips as in Section 4. For this survey, clips were not reversed. After listening to each pair, workers were asked to rate the speakers on a relative scale on competitiveness, dominance, risk-taking attitude, seniority, and trustworthiness on a seven-point Likert scale. The survey instructions and format are shown in Figure S9.

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

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

Supplementary Text

Robustness: Estimating Frequencies

I show that the estimated share of Group 1 (bimodal) clips in the sample of 6,399 verified self-recorded female lawyers is robust to various alternative specifications for estimating frequencies. The first row of Table S6 provides the baseline FMM estimation results and shows that the baseline estimated share of Group 1 clips is 0.36. Table S7 provides a selection of alternative specifications used in accordance with suggested robustness checks in the literature (e.g., Vogel et al. (2009)). Unless otherwise indicated, the baseline FMM and Praat specifications are used. Except for the first row of Table S7, all other rows in the table use a random sample of 1,000 clips from the 6,399 verified self-recorded female lawyers, though the exact number of observations may vary slightly due to failure of FMM convergence.

The first row of Table S7 shows the FMM estimation results using the residual from regressing the location of the low mode on years of experience, firm, title, and litigator fixed effects. It uses all clips in the sample of verified self-recorded female lawyers that have the demographic covariates necessary for the regression. The baseline estimated share of Group 1 clips is well within the 95% confidence intervals, showing that the FMM estimation results are robust to including demographic covariates. Table S8 further shows that the correlation between the baseline FMM estimation and residualized FMM estimation is extremely high, at 0.985.

The second row of Table S7 shows FMM estimation results with *Octave-Jump Cost* and *Voiced/Unvoiced Cost* set to 0. This specification eliminates the use of Praat's global pathfinder, such that the frequency candidate with the highest local strength in each frame, subject to the other thresholds and costs described in Section 2, is chosen as the frequency estimate for that frame without any between-frame adjustments. Without the global pathfinder, the estimated share of Group 1 clips dramatically increases from 0.36 at baseline to 0.64. Since the global pathfinder smooths the distribution of frequency estimates across frames to minimize the influence of nonhuman sounds, turning the pathfinder off will result in a cluster of frequency estimates at the pitch floor, mechanically leading to a larger share of low mode detection.

The third row of Table S7 shows FMM estimation results with *Octave-Jump Cost* increased from 0.35 at baseline to 0.5 and *Voiced/Unvoiced Cost* increased from 0.14 at baseline to 0.2. This specification increases the influence of Praat's global pathfinder, such that it more aggressively smooths the distribution of frequency estimates across frames, tending towards a less multimodal distribution of frequencies. Despite the higher hurdle to estimating a secondary mode, the baseline estimate share of Group 1 clips remains within the 95% confidence interval. Furthermore, the lower end of the 95% confidence interval exceeds 0.3, showing that the share of Group 1 clips is robust to the increased influence of the global pathfinder.

The fourth row of Table S7 shows FMM estimation results with *OctaveCost* doubled from 0.01 per octave at baseline to 0.02 per octave. Compared to the baseline, this specification favors higher-frequency candidates relatively more, making it more challenging to estimate a lower secondary mode. Despite the higher hurdle to estimating a secondary mode, the baseline estimate share of Group 1 clips remains within the 95% confidence interval. Furthermore, the lower end of the 95% confidence interval exceeds 0.3, showing that the share of Group 1 clips is robust to increased weight on higher-frequency candidates.

The fifth row of Table S7 shows FMM estimation results with the Gaussian smoothing window instead of the Hanning window. The Gaussian window has double the length of the Hanning window and provides more accurate estimates in many settings (Boersma, 1993; Boersma, 2001). The baseline estimated share of Group 1 clips remains within the 95%

confidence interval and is close to their center. Furthermore, the lower end of the 95% confidence interval exceeds 0.3, showing that the share of Group 1 clips is robust to the use of a different analysis window.

The sixth and seventh rows show results using alternative pitch floors. Changing the pitch floor from the 50 Hz baseline implies a different analysis window length. Specifically, a pitch floor or 40 Hz implies a larger analysis window of 75 milliseconds, whereas a pitch floor or 60 Hz corresponds with an analysis window of 50 milliseconds. In both cases, the results are insignificantly different from the baseline results using a window of 60 milliseconds. This also addresses a concern that Praat erroneously produces frequency estimates that are half or double the actual frequency.

Robustness: Estimating Frequency Modes

Table S9 shows aggregate model fit for a variety of mixture models using the 100 demeaned percentiles of a random subsample of 1,000 clips from all 6,399 clips in the sample of verified self-recorded female lawyers. In general, lower values of Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC) are better, but computational costs grow exponentially with g, the number of components in the mixture. The rightmost column of the table shows that $g = 5$ maximizes model fit according to BIC, and the marginal improvement from $g = 5$ onwards is below 1% according to AIC.

Table S10 shows how the estimated share of Group 1 (bimodal) and Group 2 (unimodal) clips vary with alternative specifications for estimating frequency modes. Rows 1-6 use a random subsample of 1,000 clips from these 6,399 clips, though the exact number of observations may vary slightly due to failure of FMM convergence. Row 7 uses the full sample of 6,399 clips, and uses gamma instead of normal densities with a 5-component mixture. The estimated share of bimodal clips stabilizes from Row 4 ($g = 5$) onwards, supporting Table S9 in showing that $g = 5$ maximizes the model fit.

An additional robustness check draws from Silverman (1981), who provides a theoretical framework to test if the true unobservable population density has a specific number of modes. The basic idea builds on the property that when using a Gaussian kernel to estimate a density, the number of modes is a decreasing function of the bandwidth. The standard method entails (1) locating for every $k = 1, 2, ...$ the smallest bandwidth ('critical bandwidth') that can support k modes or less, (2) generating smoothed bootstrapped samples from each critical density using a Gaussian kernel, and (3) estimating the density of each smoothed bootstrapped sample using the critical bandwidth. The proportion of samples with greater than k modes reflects the significance (i.e., p-value) of the critical bandwidth. A low p-value is evidence against the null hypothesis that the underlying density has k or fewer modes. Put differently, if all the samples indicate k or fewer modes, then the kernel density must be significantly oversmoothed to remove the appearance of mode $k + 1$. The test is seen as conservative since the bootstrapped samples are drawn from the critical density only and tends to underestimate the true number of modes. Results from this test strongly reject unimodality of the voice frequency density of female lawyers (see Table S11).

Literature on Mean Voice Frequency

Speakers with deeper voices are perceived as more attractive, dominant, mature, and honest (Imhof, 2010; O'Hair and Cody, 1987). Other studies have found that they are perceived as more truthful and empathic, and to possess greater leadership capacity (Klofstad et al., 2012; Apple et al., 1979). One study by Mayew et al. (2013) uses data from quarterly conference call recordings of public companies listed in the S&P 1500 to find that CEOs (albeit all male) with deeper voices manage larger companies.

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).

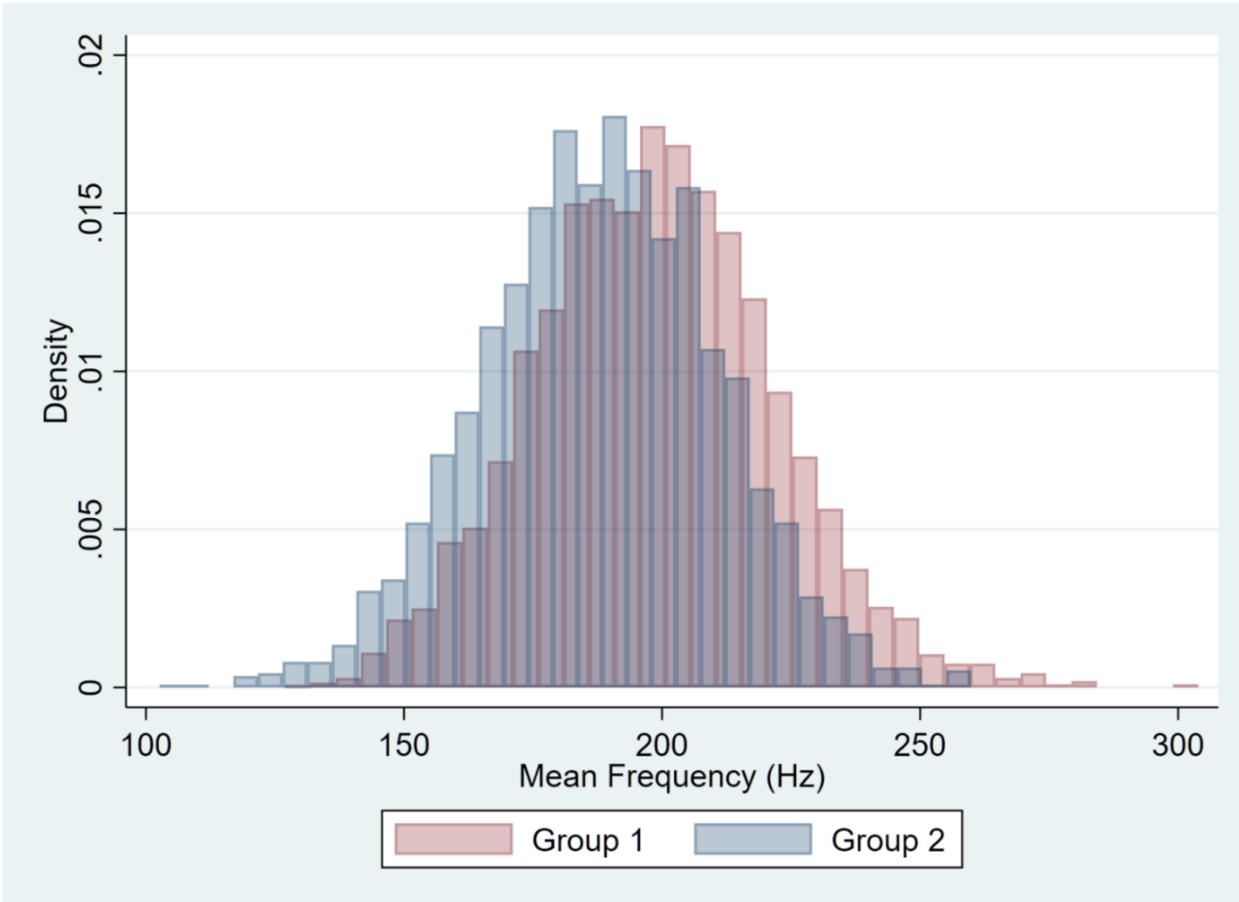


Figure S1. Histogram of mean frequencies by mode group

Notes: Figure S2 shows the histogram of mean frequencies for the 6,399 verified self-recorded female lawyer clips from the main dataset by mode group. Group 1 refers to bimodal clips, and Group 2 refers to unimodal clips. This graph only shows the distribution of *mean* frequencies *across* clips. Based on these data, there is no evidence for heterogeneity in female vocal behavior. This highlights the need to examine the distribution of frequencies *within* clips, as in Figure 1.

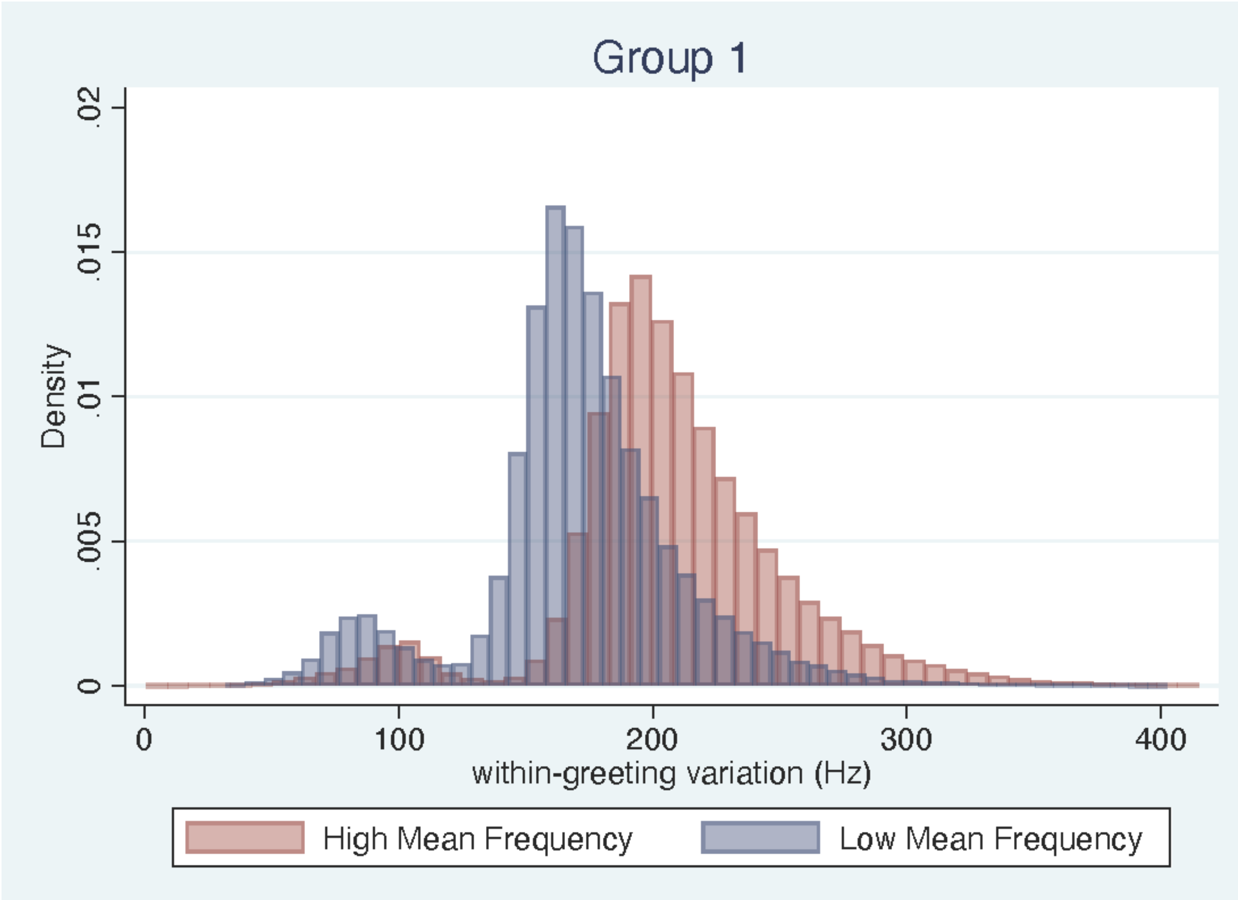
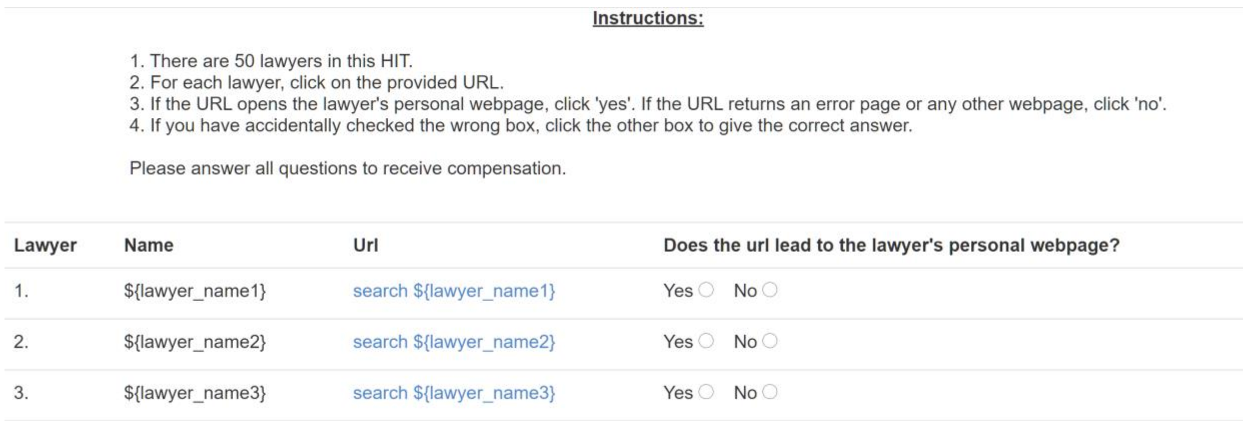


Figure S2. Histogram of Group 1 within-clip frequencies by mean voice frequency

Notes: This figure presents the within-clip frequency histograms for the 2,334 Group 1 (bimodal) clips in the sample of 6,399 verified self-recorded female lawyer clips. See main text for estimation methods. The red histogram uses Group 1 clips with mean frequencies above the Group 1 median frequency ("Group 1 high"), and the blue histogram uses Group 1 clips with mean frequencies below the Group 1 median frequency ("Group 1 low"). For Group 1 high clips, the location of the secondary mode is -109.867 Hz (std. error 0.268) away from the Group 1 high mean (206.209 Hz). For Group 1 low clips, the location of the secondary mode is -84.996 Hz (std. error 0.285) away from the Group 1 low mean (169.882 Hz). This suggests that female lawyers may mode-switch to reach a specific frequency near the unitary male mode, and that female lawyers with a higher mean voice frequency may face differential pressures to conform to the male voice frequency mode.



Instructions:

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

Please answer all questions to receive compensation.

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

Figure S3. Sample MTurk HIT for finding lawyers who switched firms

Lawyers often change firms. Our objective is to update their contact information. **We are particularly interested in their new office phone numbers.**

We are interested in all lawyers, regardless of whether they move to a law firm or another industry.

Definitions:
Personal firm webpage: This refers to URL that directly links to the lawyer's personal contact information.
Personal office phone number: This is the new office phone number that is listed alongside the lawyer's profile on his/her personal firm webpage.

Instructions:
For each lawyer, click on the provided search URL.
This will bring you to a Google page with the lawyer's name and his/her old firm name.

Case 1.
If the lawyer's new firm is on the first page of the search results, copy and paste the new firm name, the lawyer's personal firm page URL, and the lawyer's personal office phone number to the MTurk survey.

Case 2.
If the lawyer's new firm is not on the first page of the search results, do another Google search including the keyword 'LinkedIn' to find the lawyer's LinkedIn page.

From the LinkedIn page, find the lawyer's new firm. Copy and paste the new firm name to the MTurk survey.

Then search for the lawyer's personal firm webpage on Google using the new firm name you found on LinkedIn.

Case 2a.
If the lawyer's personal firm webpage is on the first page of the search results, then copy and paste the new firm name, the lawyer's personal firm webpage URL, and the lawyer's personal office phone number to the MTurk survey.

Case 2b.
If the lawyer's personal firm webpage is not on the first page of the search results, then copy and paste the lawyer's LinkedIn profile URL to the middle column of the survey.

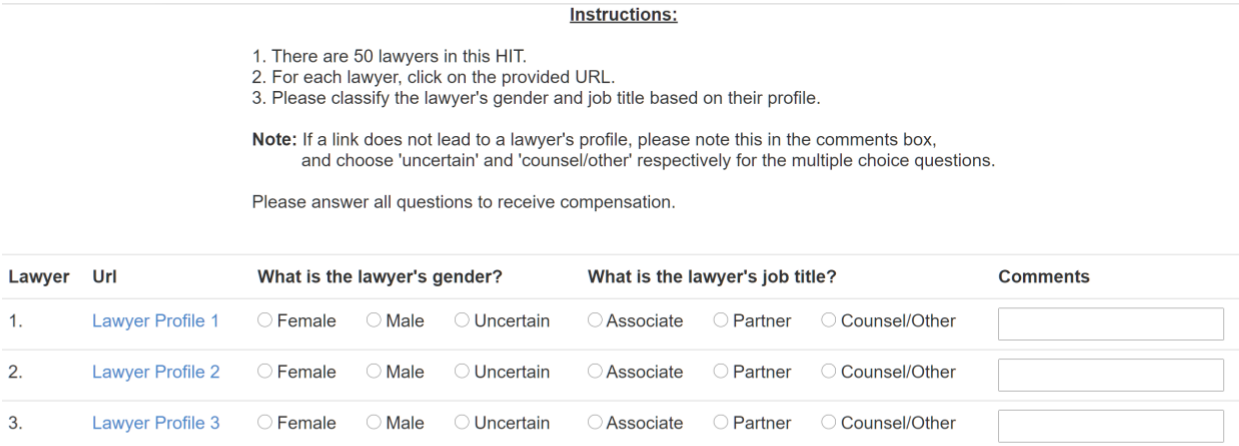
If there is any information you cannot find or if the lawyer is still at the old firm, then type NA in the relevant column.

There are 10 lawyers in this survey. This task should take about 30 minutes to complete.

Note A: Please do not use sites other than Google, LinkedIn, and the lawyer's firm website.
Note B: Please zoom out if you cannot see the whole survey.

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

Figure S4. MTurk HIT for finding new office numbers of lawyers who switched firms



Instructions:

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

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

Please answer all questions to receive compensation.

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

Figure S5. MTurk HIT for finding promoted lawyers

Instructions:

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

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

Please answer all questions to receive compensation.

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

Figure S6. MTurk HIT for classifying real estate agents' gender

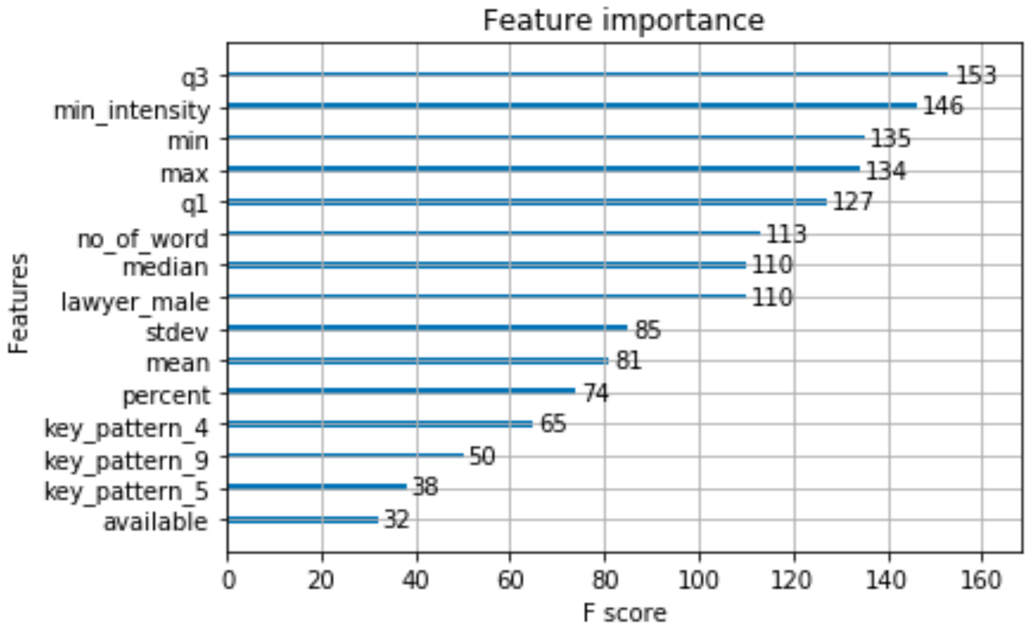


Figure S7. Ranking feature importance for machine learning classification

Notes: This figure ranks the fifteen most important attributes used in the XGBoost model for identifying voicemails that were self-recorded by lawyers. For each decision tree, the importance of each attribute is calculated as the amount by which 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.

Requirements

- Please note that you can do **at most 1 HIT from this batch**. You will not be compensated for additional HITs.
- A quiet environment and noise-cancelling headphones are essential to succeed in this task.
- This survey will not function properly on Safari. Please use another browser, such as Mozilla Firefox or Google Chrome.
- A high-speed Internet connection is strongly recommended.

Overview

- This survey takes about 30 minutes to complete, with a time limit of 3 hours.
- There are 5 sections in this survey, each with 10 questions.
- In each question, you are presented with a pair of 3-second audio recordings.
- Both clips have a similar mean pitch, but one clip has a more bimodal distribution of pitch ("bimodal") relative to the other clip ("unimodal").
- Your task is to identify the bimodal clip.

Bonus Compensation

- If your mean success rate across sections exceeds 0.5, then you will earn bonus compensation equal to:
    - $(maximum success rate + mean success rate) x 1.25.
- For example, if you correctly answer 8 of 10 questions in a section, then your success rate is 0.8. If this is also your best score across all five sections, then this is your maximum success rate.
- If you correctly answer 6 of 10 questions in each of the other 4 sections, then your mean success rate is (0.6 x 4 + 0.8 x 1)/5 = 0.64.
- Your bonus is then $(0.8 + 0.64) x 1.25 = $1.80.
- On average, previous survey-takers have been able to improve their success rate over time, obtaining a mean success rate of approximately 0.6.

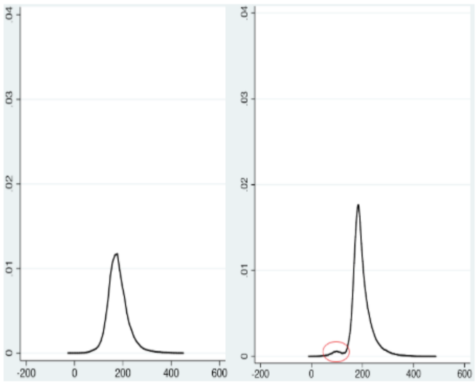
BEGIN

Instructions:
If necessary, please zoom out to see the full survey.

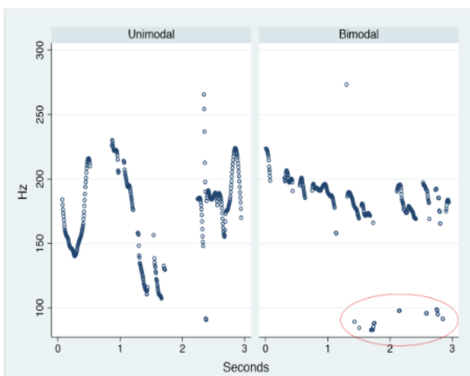
Your task is to classify audio recordings based on their pitch properties.
To preserve the pitch properties and remove the influence of verbal content, the clips are played from finish to start.

There are two recordings per question. Listen to both recordings as many times as necessary.
In each question, both clips have a very similar mean pitch (less than 1 Hz apart), but one has a minor mode with a lower pitch.
You are asked to identify which of the two clips exhibits a more bimodal distribution of pitch.

To illustrate, the figure below shows a unimodal (left) and bimodal (right) distribution of pitch.
Both distributions have a mean pitch of about 200 Hz, but only one has a minor mode at about 100 Hz (circled in red).



Likewise, the following figure shows an example of how pitch varies over time. The bimodal clip has a cluster of points near the 100 Hz mark, circled in red.



Before you begin the task, listen to the following practice clips and view the corresponding videos to get a sense of the differences between the unimodal and bimodal pitch.
Previous survey-takers have described the bimodal recording as having a "scratchier" sound than the unimodal recording.
Please note that only audio recordings will be available once you begin the task.

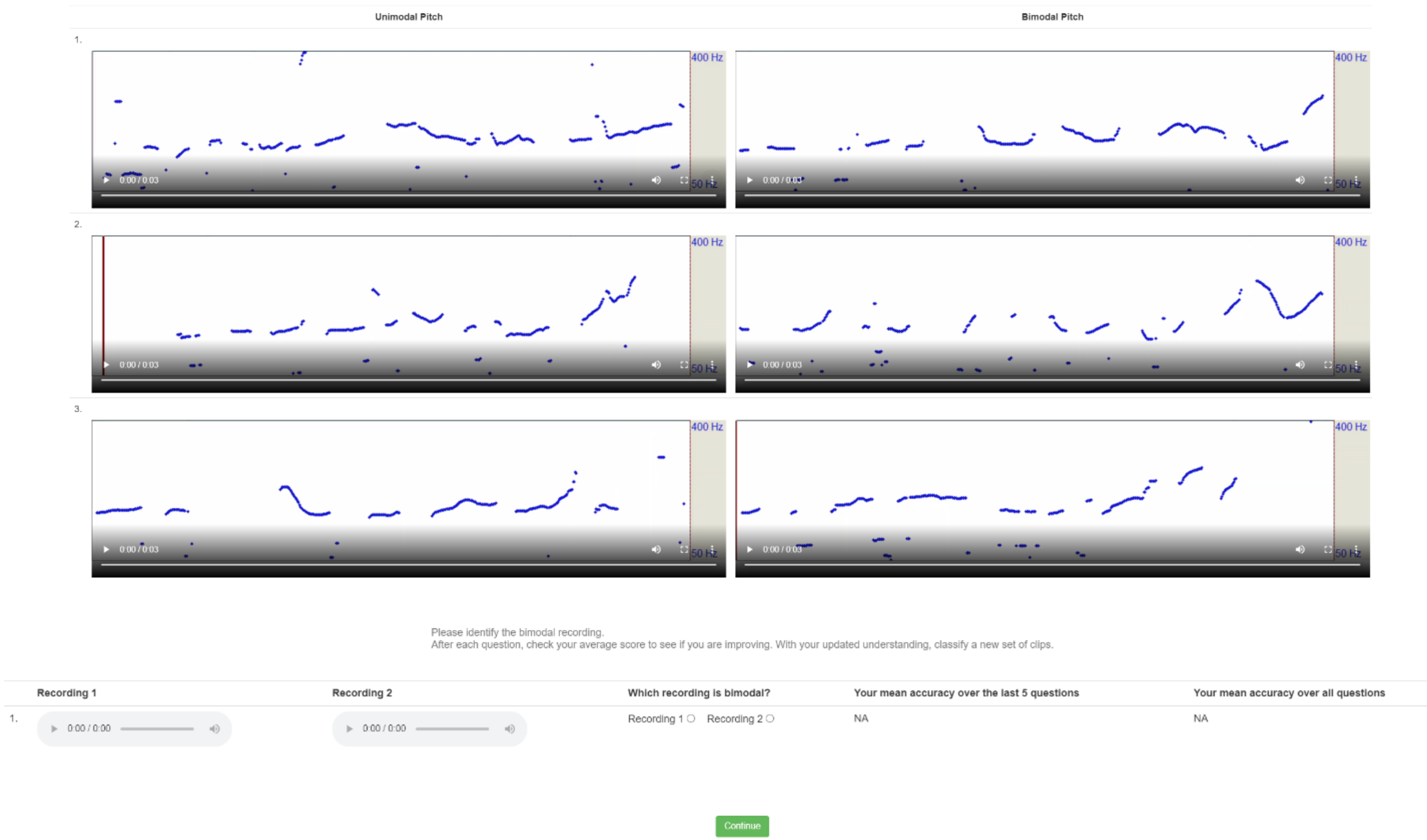
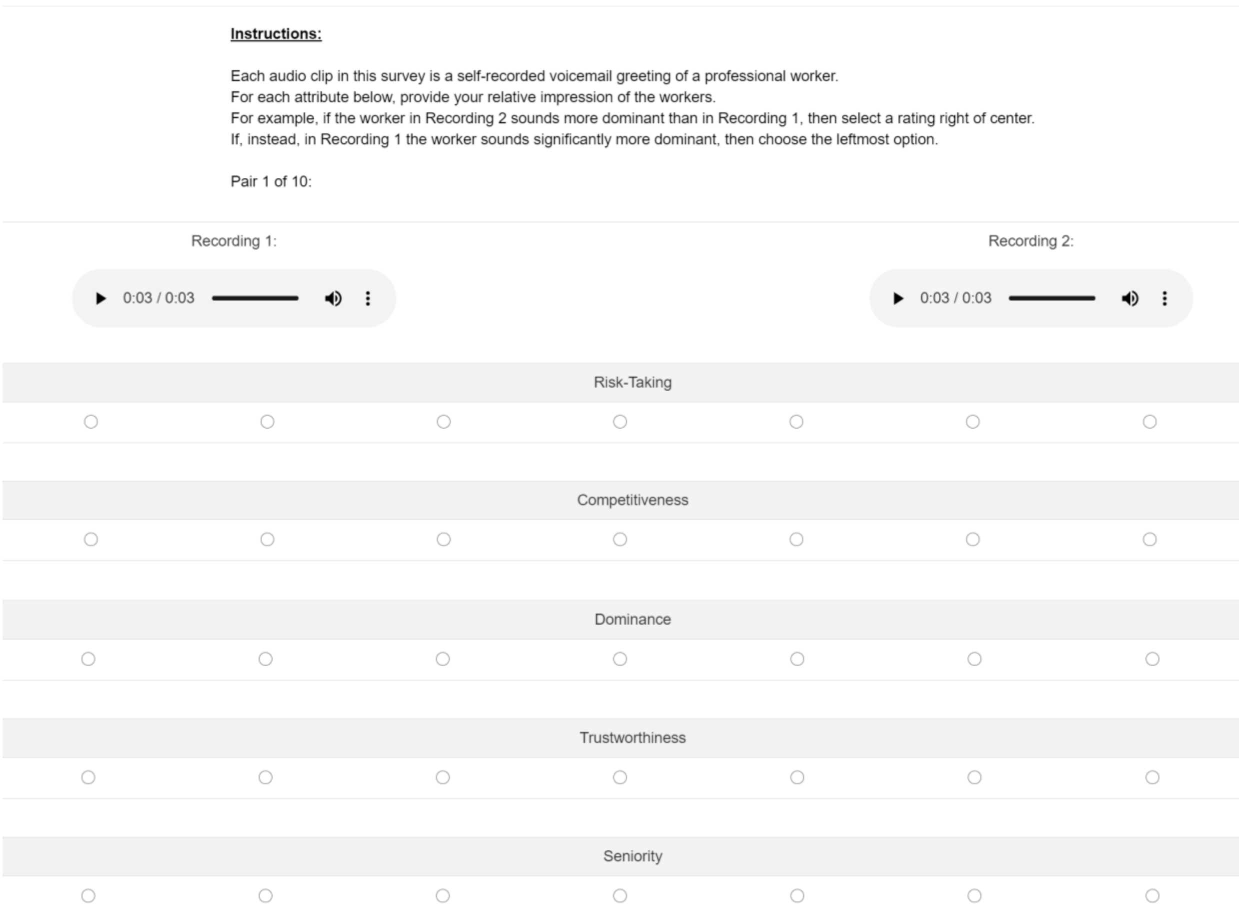


Figure S8. MTurk HIT for perception of bimodality in voicemail greetings

Notes: This figure (pages 22-24) shows the MTurk HIT instructions and format for the original survey to test whether MTurk workers could perceive bimodality in lawyers' voicemail recordings. The follow-up survey had 3 sections instead of the original 5 and double the bonus compensation of the original survey, but was otherwise identical. See Section 4 for details on the survey design.



Instructions:

Each audio clip in this survey is a self-recorded voicemail greeting of a professional worker.
For each attribute below, provide your relative impression of the workers.
For example, if the worker in Recording 2 sounds more dominant than in Recording 1, then select a rating right of center.
If, instead, in Recording 1 the worker sounds significantly more dominant, then choose the leftmost option.

Pair 1 of 10:

Recording 1:

0:03 / 0:03

Recording 2:

0:03 / 0:03

Risk-Taking

Competitiveness

Dominance

Trustworthiness

Seniority

Figure S9. MTurk HIT for relative rating of unimodal and bimodal clips

Table S1. List of law firms used in the main analysis

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

Table S2. Most common words in voicemail greetings

	Not Self-Recorded	Self-Recorded	*t*-stat
reached	0.04	0.24	-14.03
office	0.03	0.05	-2.20
take	0.01	0.08	-8.12
voicemail	0.01	0.05	-4.75
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
away	0.00	0.03	-5.82
hi	0.01	0.19	-15.52
available	0.38	0.07	18.92
sorry	0.25	0.07	12.03
unable	0.00	0.02	-4.64
often	0.00	0.01	-2.29
now	0.01	0.08	-8.01
currently	0.00	0.03	-5.09
get	0.00	0.03	-4.16
recorded	0.02	0.00	5.16
time	0.00	0.01	-1.89
return	0.00	0.01	-2.95
method	0.03	0.00	4.61
recall	0.02	0.00	4.09
desk	0.00	0.02	-4.50
preach	0.00	0.01	-4.15
name	0.01	0.03	-4.03
phone	0.00	0.03	-5.24
thank	0.01	0.01	-0.64
one	0.02	0.01	0.74
think	0.01	0.01	-0.86
got	0.01	0.01	-0.23
freak	0.00	0.01	-3.76
missed	0.00	0.02	-4.39
mailbox	0.01	0.01	0.67
either	0.00	0.01	-2.42
okay	0.01	0.00	0.42

forwarded	0.01	0.00	2.84
messaging	0.01	0.00	2.46
system	0.01	0.00	2.76
met	0.01	0.00	2.58
cared	0.00	0.00	0.63
god	0.00	0.00	-0.74
great	0.00	0.01	-0.90
another	0.00	0.01	-2.08
map	0.00	0.00	-2.24
firm	0.00	0.01	-3.02
eight	0.00	0.00	0.09
table	0.00	0.00	-0.52
dc	0.00	0.00	-1.73
Total words	3.40	5.41	-33.51

Notes: This table shows the most frequent words and the number of words per clip for lawyer voicemails in the main dataset. These attributes were used as machine learning predictors for identifying voicemails that were self-recorded by lawyers.

Table S3. Frequently-occurring phrase patterns in voicemail greetings

Variable Name	Significant Phrase
Key pattern 1	This is / This
Key pattern 2	You have reached / You reached/ You reach/ YMy reached
Key pattern 3	office of/ desk of/voice mail of/ voicemail of/mail
Key pattern 4	I am/am/my/mine
Key pattern 5	Away from/on the phone/away/phone
Key pattern 6	Record message/record a message
Key pattern 7	leave a message/leave message/left a message/left message
Key pattern 8	Missed yMy call/miss you call/missed you call
Key pattern 9	is
Key pattern 10	you/your/yours
Key pattern 11	Call/calling/called/k
Key pattern 12	Thank/thanks/thank you

Notes: This table shows the 12 frequently-occurring phrase patterns in lawyer voicemails in the main dataset. These attributes were used as machine learning predictors for identifying voicemails that were self-recorded by lawyers.

Table S4. Acoustic variables used as machine learning predictors

Variable Name	Description
max	The maximum frequency of the clip
min	The minimum frequency of the clip
$median$	The median frequency of the clip
$mean$	The mean frequency of the clip
$stdev$	The standard deviation of the frequency of the clip
min_inten	The minimum value of intensity of the clip
qi	The ith quantile frequency of the clip, $1 \leq i \leq 100$
int_i	The intensity of ith interval of the clip. Each clip is divided into 225 intervals with equal length. $1 \leq i \leq 225$

Notes: This table shows the acoustic features of lawyer voicemails in the main dataset that were used as machine learning predictors for identifying voicemails that were self-recorded by lawyers.

Table S5. Comparison of machine learning models

Model	Accuracy Ratio
Random Forest	86.66%
Support Vector Machine	87.27%
K-nearest Neighbours	85.04%
XGBoost	93.52%

Notes: This table compares the performance of four machine learning models in identifying voicemails that were self-recorded by lawyers. The table reports the predictive accuracy of the models on the testing sample using a probability threshold of 0.5.

Table S6. FMM results for Figure 3b

Sample	Low Mode (Share)	Low Mode (Location)	Primary Mode (Location)	Observations
All Verified Lawyers	0.360 (0.007)	86.139 (0.381)	162.481 (0.423)	6,399
Partners	0.385 (0.036)	87.114 (1.819)	166.485 (2.009)	196
Counsel/Other	0.414 (0.042)	87.794 (1.910)	174.106 (2.519)	137
Switchers (Previous Firm)	0.338 (0.034)	88.704 (1.709)	163.450 (1.935)	198
Switchers (New Firm)	0.380 (0.036)	86.103 (1.682)	158.741 (2.301)	198
SCOTUS (beginning)	0.329 (0.044)	80.136 (3.932)	177.978 (3.011)	129
SCOTUS (middle)	0.380 (0.046)	87.720 (3.436)	175.494 (2.836)	129
SCOTUS (end)	0.410 (0.047)	85.402 (2.969)	177.180 (3.272)	129
Assistants (Female Lawyer)	0.387 (0.0336)	82.851 (1.724)	157.590 (1.892)	237
Assistants (Male Lawyer)	0.255 (0.028)	84.435 (3.334)	162.179 (1.616)	412
REMAX Residential	0.214 (0.027)	87.717 (2.899)	171.143 (1.880)	337
REMAX Commercial	0.181 (0.042)	84.948 (5.880)	164.107 (2.819)	159

Notes: Refer to main text for descriptions of samples. Delta method standard errors in parentheses.

Table S7. Robustness of unobserved heterogeneity to frequency measures

Frequency Estimates	Group	95% C.I. (Share)	95% C.I. (Location)	Group	95% C.I. (Share)	95% C.I. (Location)	Obs.
Residualized Baseline	1	(0.347, 0.380)	(-48.243, -46.326)	2	(0.620, 0.653)	(25.943, 28.044)	4,682
No Pathfinder	1	(0.611, 0.674)	(80.938, 83.021)	2	(0.326, 0.389)	(158.863, 164.568)	999
Stronger Pathfinder	1	(0.313, 0.378)	(83.361, 87.649)	2	(0.622, 0.687)	(160.376, 164.321)	998
Double OctaveCost	1	(0.309, 0.374)	(83.709, 87.763)	2	(0.626, 0.691)	(160.617, 164.596)	997
Gaussian Window	1	(0.330, 0.395)	(84.465, 88.315)	2	(0.605, 0.670)	(160.947, 164.894)	998
60 Hz Floor	1	(0.340, 0.408)	(84.672, 87.711)	2	(0.592, 0.660)	(157.858, 162.423)	997
40 Hz Floor	1	(0.297, 0.362)	(84.189, 88.385)	2	(0.638, 0.703)	(161.029, 164.884)	988

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

Table S8. Robustness: Low mode prediction

Correlation matrix	Fixed effects model	Base model
Fixed effects model	1	
Base model	0.985	1

Notes: This table shows the correlation between the baseline FMM model and the frequency estimates using the residual from regressing the location of the low mode on years of experience, firm, title, and litigator fixed effects. $n = 4{,}682$. See first row of Table S7 for more information on the fixed effects model used.

Table S9. FMM optimal number of components

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

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

Table S10. Robustness of unobserved heterogeneity to low mode measures

Individual FMM Comp.	Group	95% C.I. (Share)	95% C.I. (Location)	Group	95% C.I. (Share)	95% C.I. (Location)	Obs.
2	1	(0.085, 0.124)	(84.306, 90.325)	2	(0.876, 0.915)	(179.291, 182.219)	1,000
3	1	(0.275, 0.335)	(84.854, 88.660)	2	(0.665, 0.725)	(169.137, 172.829)	999
4	1	(0.324, 0.389)	(84.979, 88.612)	2	(0.611, 0.676)	(163.034, 167.051)	998
5	1	(0.348, 0.416)	(84.535, 88.411)	2	(0.584, 0.652)	(160.970, 165.396)	1,000
6	1	(0.371, 0.439)	(84.177, 87.852)	2	(0.561, 0.629)	(159.477, 163.799)	990
7	1	(0.365, 0.436)	(84.496, 88.500)	2	(0.564, 0.635)	(157.795, 162.334)	989
5 (Gamma)	1	(0.371, 0.398)	(87.362, 89.116)	2	(0.602, 0.629)	(163.494, 165.054)	6,399

Notes: This table shows how the estimated share of Group 1 (bimodal) and Group 2 (unimodal) clips vary with alternative specifications for estimating the clip-level frequency modes. Results are from a 2-component FMM applied to the different individual low mode estimates. The number of observations column refers to the clip-level FMM estimation of the 1,000 female lawyers described above. Missing observations are due to clip-level convergence failures. The bottom row presents results from the full sample of verified female lawyers, substituting normal with gamma densities in a 5-component mixture. The results are similar for all but the 2 and 3 component models. For these, the mode shares were significantly skewed toward the primary mode relative to the other results. See Supplementary Text for details.

Table S11. Silverman test for number of modes

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

Notes: The table shows the results of the Silverman test against the null hypothesis that the underlying frequency density has m or fewer modes. This test used demeaned data from 1,000 randomly selected clips from the main dataset that were self-recorded by female lawyers. For each additional mode, the test is performed by creating 50 perturbed samples of the 100,000 frequency percentiles using the critical bandwidth identified with precision of 0.01 Hz or lower. The *p-value* is the proportion of samples that produced more modes than stated by the null. The Silverman test is seen as conservative since the bootstrapped samples are drawn from the critical density only and tends to underestimate the true number of modes. The results strongly reject the null hypothesis that the frequency density has a single mode (*p-value* = 0), but cannot reject that the number of modes is 2 or more. The mode location refers to the distance from one's mean frequency, estimated using 40 averaged shifted histograms (ASH-WARPing). Modes 3 and above are more than 100 Hz from the mean.