Gender and Tone in Recorded Economics Presentations: Audio Analysis with Machine Learning[*]

Amy Handlan[†] and Haoyu Sheng[†]

January 1, 2023

Click here for most recent draft

Abstract

This paper develops a replicable and scalable method for analyzing tone in economics seminars to study the relationship between speaker gender, age, and tone in both static and dynamic settings. We train a deep convolutional neural network on public audio data from the computer science literature to impute labels for gender, age, and multiple tones, like happy, neutral, angry, and fearful. We apply our trained algorithm to a topically representative sample of presentations from the 2022 NBER Summer Institute. Overall, our results highlight systematic differences in presentation dynamics by gender, field, and format. We find that female economists are more likely to speak in a positive tone and are less likely to be spoken to in a positive tone, even by other women. We find that male economists are significantly more likely to sound angry or stern compared to female economists. Despite finding that female and male presenters receive a similar number of interruptions and questions, we find slightly longer interruptions for female presenters. Our trained algorithm can be applied to other economics presentation recordings for continued analysis of seminar dynamics.

Keywords: machine learning, audio analysis, gender, economics profession

JEL Codes: A1, C8, C45, J7

[*] First, we thank NBER for providing data for this project. Without Jim Poterba, Alterra Milone, Kristen Kenny, Debby Nicholson, and others on the NBER team, this project would not be possible. We thank Joanna Lahey for her discussion and we thank Anna Aizer, John Friedman, Peter Hull, Soonwoo Kwon, and Emily Moschini for their helpful conversations and suggestions. We thank Frank Chiu for his research assistance.

[†] Department of Economics, Brown University. Handlan email: amy_handlan@brown.edu. Sheng email: haoyu_sheng@brown.edu

1 Introduction

Seminars and presentations are critical components in the development and dissemination of one's research. Presentations allow researchers to collect feedback, advertise projects, and collect signals on how their work will be received by others. Recent work by Dupas et al. (2021) quantifies that female and male economists have systematically different experiences in presentations with different likelihoods to be interrupted and for those interruptions to have more patronizing or hostile tones. A gender gap at the presentation stage of the research pipeline can factor into additional differences downstream through publishing (Card et al., 2020; Hengel, 2022; Hengel and Moon, 2022), attribution of credit in coauthored projects (Hengel and Phythian-Adams, 2022; Sarsons, 2017; Sarsons et al., 2021), promotion (Bosquet et al., 2019; Ginther and Kahn, 2021; Lundberg, 2022), recognition and awards (Card et al., 2022), and more. Accordingly, measuring differences in the quality of presentations is an important element for understanding the status of women in the economics profession.

However, quantifying the quality of a presentation is not trivial. Should we measure the presentation's content or delivery, study the behavior of the presenter or the audience, or do both? Ultimately, a presentation's quality depends on both the presenter—what content they choose to share and the manner in which they deliver that information—and the audience—in how they engage with the presenter. For now, we focus on measuring variations in how people speak with variations in tone throughout a presentation. Measures of tone over time allow researchers to compare communication patterns across different speakers (presenters, discussants, and speaking audience members) from different fields (as content quality will be field specific). Accordingly, conditional on the presentation content being of similar quality, if quantitative measures of how people speak are systematically different for female and male economists, then this is indicative of gender bias in presentations.

In this paper, we develop a machine learning approach that allows for both scalable and time-consistent measures of the tone of speakers in recorded economic presentations. We apply our algorithm to recordings of presentations at the National Bureau of Economic

Research's Summer Institute conference (hereafter, NBER SI), a prestigious multi-field economics conference. We then use the imputed tone measures to study the dynamics of tone within a speaker and between speakers during presentations. The data are aggregated and anonymized to abstract from individual sessions and speakers and to focus on broader patterns. We document which tones are more or less likely to be used by female versus male and junior versus senior economists in seminars after controlling for aggregated fields (macroeconomics and microeconomics). To study the seminar dynamics, we quantify changes in tone over time as the discussion switches between the presenter and audience members. From this, we estimate how past presenter-audience interaction changes future presenter tones. We follow the footsteps of Dupas et al. (2021) in seeking to quantify and study gender and economics presentation culture.[1] We find that both male and female economists are less likely likely to sound positive when responding to a female speaker. Finally, we outline best practices for standardizing audio data and training robust machine learning models to impute labels for economics research.

To perform this imputation, we trained a convolutional neural network on public datasets of audio clips with clearly defined emotion labels and speaker metadata that allows us to impute labels from raw audio recordings. Since March 2020, many economics seminars and conferences record presentations and publicly broadcast them, at least temporarily, online. The NBER SI is one such conference that has started providing a hybrid option and streaming the presentations online.[2] We were able to work with recordings for 479 paper presentations from the 2022 conference and input the recordings into our algorithm to produce anonymized measures of tone, gender, and age of all speakers—including presenters, discussants, and speaking audience members.

There are three innovations in our machine learning approach relative to a human-coding

[1] See the related literature section for a more detailed discussion about how we relate to Dupas et al. (2021); however, our novel machine learning approach allows us to provide both new empirical findings and a methodological contribution.

[2] Participants in the NBER SI are notified in their welcome letter that the conference will be streamed, and they were also notified that the recordings would be used in research. Please see Section 2 for more information.

approach. First, the machine learning approach allows for a scalable, hands-off approach to labeling tone. With two researchers, we were able to measure tone for our entire sample without sitting through all presentations, some of which occur simultaneously. Second, our measures of tone include probability scores across six categories (happy, sad, angry, neutral, sad, and disgust) rather than across discrete categorical labels, meaning we can analyze audio segments as being more or less happy. Third, our measures of tone are high frequency in that we evaluate tone at five-second intervals, allowing us to study tone variation at different levels of analysis ranging from five-second intervals, to periods of uninterrupted speech, to the presentation level. It also allows us to systematically assign tone labels to every utterance by all speakers, including presenters, discussants, and audience members. These features of our tone measures allow us to conduct novel empirical analysis on the dynamics of tone in economics seminars. Furthermore, we are making our audio-tone imputation algorithm publicly available as another contribution to the literature and the profession for continued study of presentation dynamics in different settings but with the same metrics.

The main findings of our paper can be broken down into more detail by two levels of analysis: within speaker and between speakers. We first produce summary statistics and cross-sectional regression analysis about the gender, tone, and age of presenters and speaking audience members in presentations. We then control for the research field (macroeconomics or microeconomics), the presentation format (regular seminar or seminar with discussant), and the overall share of female speakers in the presentation. Subsequently, we leverage the high-frequency timing of our data to control for the lagged tone of speakers and to study changes in tones between speakers.

We first find that the representation of female economists at the NBER SI has increased. The representation of female presenters has increased dramatically to an almost equal split compared to Chari and Goldsmith-Pinkham (2017). However, there is still evidence of same-gender sorting: women in the audience are more likely to ask female presenters questions than male presenters, and female presenters are more likely to be assigned female discussants. This effect remains even after controlling for the overall share of female speakers in a talk, which we use as a proxy for female-dominated sub-fields in economics. One important point of clarification is that our measure of audience participation slightly differs from the literature, in that ours is based on speaking rather than on attendance.

Regarding interruptions, we find that there are similar number of interruptions for both male and female presenters. This differs from Dupas et al. (2021), a closely related paper in the literature studying economics seminar dynamics with hand-coded data, who find that female presenters are interrupted more than their male counterparts and discussant format presentations have fewer questions. Nevertheless, we do find some results that are consistent with Dupas et al. (2021). For instance, we find that interruptions for female presenters last longer than those for male presenters and that macroeconomics talks have more interruptions and questions relative to microeconomics talks. We also find that there are similar amounts of audience member participation in both regular format and discussant format talks at the 2022 NBER SI.

In addition to interruption counts and duration, we study variation in tone to further analyze differences in seminar dynamics. Cross-sectionally, we find gender differences in tone within speakers. On average, female speakers are more likely to sound positive or happy, while male speakers are more likely to sound negative and serious or angry. This holds whether we consider only presenters or only speaking audience members. Furthermore, when we consider how speakers may change their tone over time, we find that tone is highly persistent for both men and women. That is, if you sound happy and are uninterrupted, you are likely to continue sounding happy.

This persistence is important when studying the effect of interruptions and dynamics between speakers. Overall, we find that speakers sound more negative when responding to women, whether the person is an audience member asking a question to a female speaker or a presenter responding to a female audience member. When we look at the interaction between presenter gender and audience gender, we find that female speakers respond more negatively to other women compared to how they respond to men. The gap in tone responding to men versus women is larger for female speakers than for male speakers.

Here we think there are two effects at play: first, a common theory across social sciences highlights that societal norms and heuristics lead to people having higher standards and expectations for women while simultaneously having negative beliefs about female ability (Chevalier et al., 2021; Hengel, 2022; Sarsons and Xu, 2021). This is one way to generate the negative response to women and the negative interaction effect for women responding to other women. Beyond the direct effects, there is also an indirect channel. Women may speak more positively to men in an attempt to offset a larger negative bias from men compared to women. This channel is similar to the finding in Hengel (2022), who argues that female economists write better papers before submission to journals in anticipation of discrimination. In our current methodology, we assign weights to the two channels and think both factor into current tone dynamics.

Our findings also speak to how speaker age, field, and seminar type affect tones and interactions. We find that speakers who sound older tend to sound happier and less serious. Speakers in macroeconomics seminars are less likely to sound happy and serious and are more likely to sound sad and neutral, and speakers in seminars without discussants are more likely to sound serious. Presenters who sound older are less likely to respond happily to audience members, and audience members who sound older are more likely to respond happily to presenters in seminars with discussants. Presenters in macroeconomics seminars with no discussant experience more large positive changes in how likely they are to sound happy in response to an audience member's question and even more large negative shifts. These findings are consistent with what one might expect with regards to speaker seniority and seminar culture across fields.

1.1 Related Literature

Our paper contributes to two broad strands of literature, with the first being the literature on the applications of machine learning methods with audio data. Within this literature, there is a vast interdisciplinary array of work that develops machine learning methods for audio analysis. In relation to this literature, we follow recommendations from computer science textbooks, such as Camastra and Vinciarelli (2015) and Hastie et al. (2009), and commonly cited papers using convolutional neural networks for audio classification tasks (Badshah et al., 2017; Issa et al., 2020; Lim et al., 2016; Zeng et al., 2019; Zhang et al., 2018; Zisad et al., 2020). In our paper we use a supervised learning and ensemble learning algorithm to be able to impute gender, tone, and age labels from raw audio recordings.

There are many papers in linguistics and computer science highlighting that machine learning algorithms can learn unintended biases from training data. One example is Story and Bunton (2015), which shows that working with the audio data of children's speech requires a separate model from adults due to their different vocal range and fluctuations. In our paper, we show that training a joint model to impute tone from both female and male speech will lead to a higher misclassification of tones for the same reason. Accordingly, we develop separate, parallel algorithms for imputing tone for men and women as a way to account for gender fixed effects of speech since gender differentiation has been shown to increase speech emotion recognition accuracy for naive Bayes classifiers (Vogt and André, 2006).[3]

Outside of computer science, applications of audio analysis are common in medical and biological settings. For example, biologists use machine learning to classify different species by their sounds (Kahl et al., 2022). In healthcare contexts, practitioners and researchers have studied how to use classification on audio data to aid in diagnosis of different diseases, such as respiratory conditions (Aykanat et al., 2017; Xia et al., 2022).

[3] We acknowledge that some economists do not identify as strictly male or female, but the training data equates sex with gender and lacks non-binary data. Due to these data limitations, we focus on the differences between male and female speakers.

In economics, the number of papers using machine learning approaches for audio analysis is increasing. There are a growing number of papers studying the voices of policymakers: from the tone of Federal Reserve chairs in speeches that affect monetary policy (Alexopoulos et al., 2022; Bisbee et al., 2022; Gorodnichenko et al., 2021) to distinguishing fake from real presidential speeches (Alves et al., 2019). In labor economics, researchers have investigated how audio analysis during job interviews can aid recruiters in screening candidates (Liem et al., 2018; Naim et al., 2015; Nguyen et al., 2014; Teodorescu et al., 2022). In children and education economics, researchers have used audio recordings of teacher and student behavior to assess classroom climate and learning (James et al., 2018). Given the increased application of these methods, one of the contributions of our paper is to provide some guidelines for imputing labels from audio recordings, the main way economists will use this tool. Our contribution here is not in developing an entirely new method but rather in aiding other economists in appropriately using the available methods.

The second strand of literature we contribute to studies the status of women in the economics profession and, more specifically, gender differences in presentations. [4] The Committee on the Status of Women in the Economics Profession (CSWEP) releases many newsletters that cover how the representation of women in economics has evolved over time as well as a yearly session at the ASSA conference highlighting papers studying this same topic. Through newsletters issued by the CSWEP, published papers, and ongoing projects, this literature has documented systematic differences between female and male economists at every stage of their academic career, including field selection (Avilova and Goldin, 2018; Dolado et al., 2011; Sierminska and Oaxaca, 2021), hiring and promotion (Bosquet et al., 2019; CSWEP Committee on the Status of Women in the Economics Profession, 2022; Ginther and Kahn, 2021; Lundberg, 2022), and publication and recognition (Card et al., 2022; Hengel and Phythian-Adams, 2022; Sarsons, 2017; Sarsons et al., 2021).[5]

[4] There is a growing and important literature studying discrimination and the status of underrepresented minorities (URM) in economics. Given that we cannot control for race or nationality, we cannot produce measures for presentations by URM economists.

[5] Abrevaya and Hamermesh (2012) find no gender differences in referee treatments of papers.

Among these many dimensions, our work most closely relates to the analysis of gender differences in economics presentations, which three other papers study—Chari and Goldsmith-Pinkham (2017), Doleac et al. (2021), and Dupas et al. (2021). Chari and Goldsmith-Pinkham (2017) document the representation of women at the NBER SI over time, from 2005 to 2016, through authorship of accepted papers and as session organizers. They find that there is an increasing share of female-authored papers at the NBER SI over time, starting with about 30 percent of papers having one female author in 2001–2004 and 40 percent of papers having at least one female author by the 2016 NBER SI. They also find that papers with a female author are more likely to be in a session with a female organizer and in a microeconomics field. One limitation of using NBER SI programs is that Chari and Goldsmith-Pinkham (2017) cannot identify the presenting author.

Looking specifically at presenting authors but outside the NBER SI, Doleac et al. (2021) find a similar increase in the representation of women as presenters in invited seminars at top institutions from 2014-2019. They document that the share of female seminar speakers increases from about 20 percent to over 30 percent by the end of their sample.[6] Encompassing both domains, Dupas et al. (2021) analyze presenters both at internal department seminars and at the 2019 NBER SI. They find that 30 percent of presenters are women at the NBER SI and at internal department seminars. They also consider recruiting seminars and find an almost equal gender split for job market candidates presenting flyout seminars. Similar to Chari and Goldsmith-Pinkham (2017), Dupas et al. (2021) find that microeconomics seminars have a greater representation of female economists compared to macroeconomics seminars. Measuring presenter representation is important to understanding the status of women in the economics profession, but Dupas et al. (2021) take it a step further and also quantify experiences within seminars.

Not all seminars are created equal, and deciding how to quantitatively measure dynamics within a seminar is crucial to documenting, understanding, and addressing those differences.

[6] In their paper, they conduct analysis by gender and underrepresented minority status, but here we report the aggregated result for women.

Dupas et al. (2021) take on this challenge and collect a large dataset of seminar dynamics using almost 100 hand coders to cover all the presentations in their sample. In addition to presenter characteristics, they produce a set of positive measures—audience size and composition, and the quantity and type of interruptions—and normative measures—the tone of the interruption.[7] They find that female presenters experience more interruptions than their male counterparts in job talks, in regular department seminars, and at the NBER SI. However, it is unclear if more interruptions is inherently good or bad. Accordingly, they also ask coders to label the tone of interruptions from a set of five tone labels: supportive, patronizing, disruptive, demeaning, and hostile. For department seminars, coders assign tone for 15 percent of interruptions in regular seminars and 8 percent of interruptions in job talks. When matching this with presenter characteristics, they find that female presenters are more likely to receive patronizing or hostile interruptions.

The quantitative measures from both positive and subjective elements in seminars have greatly influenced discussions of how to improve the seminar culture in economics. Along with other works, Dupas et al. (2021) have influenced the AEA's guidelines for ensuring constructive seminars and conferences[8]. It is clear that the profession wants to measure tone dynamics in seminars to gauge the effectiveness of promoting diversity and inclusivity.

Looking forward, we see there is room to expand on the important and pioneering work by Dupas et al. (2021) through introducing a machine learning algorithm for coding seminars. Training such an algorithm, applying it to economic seminars, and providing it for future analysis is a key methodological contribution of our paper. Switching from hand coding to an automated coding approach has advantages in terms of time and consistency. Hand coding is costly in terms of the time needed for training, execution, management, and validation; Dupas et al. (2021) detail these steps in their paper. Meanwhile, with a machine learning

[7] Interruptions in Dupas et al. (2021) are coded with the interruption type (comment, criticism, suggestion, clarification, or follow-up), the characteristics of the questioner, and how the presenter responded (i.e., did they answer or defer the question).

[8] The guidelines can be found at https://www.aeaweb.org/ resources/ best-practices/ conducting-research.

algorithm, we scale up the labeling of speech and interruptions with only two researchers.[9] Another advantage of using algorithmic tone assignment is that the same measuring tool is used across presentations and can be consistent over years. The societal norms that influence human coders will inherently change over time and endogenously respond to updated information on seminar experiences by gender.

However, there are trade-offs when using our machine learning approach relative to human coders. For example, the tones we consider, such as happy, sad, and angry, are fairly generic and are not tone labels specific to asking questions—like the tone of patronizing from Dupas et al. (2021). The tone labels we use are limited by the public training data that we use. More generally, with machine learning on audio data, we can only measure emotion and tone independent of content, whereas human coders can incorporate both. That is, we measure "how people said something," not "what people said and how they said it." A final trade-off comes from seminar access: algorithms require recordings of presentations for analysis, while human coders do not. Nevertheless, recordings would likely be beneficial to both computer and human coders in that they would allow for replicability. As this literature continues to grow, it is likely that a combined approach will allow for more nuance and provide more precision to the analysis. We leave this to future research.

The rest of the paper proceeds as follows. Section 2 details our data sources and how we produce the tone, gender, and age labels from audio data. Section 3 covers our summary statistics and empirical analysis. Here we look at cross-sectional regressions and also leverage the timing dimension of our data to look at tone within speaker and between speakers. In Section 4, we discuss the broader implications of our approaches. Section 5 concludes.

[9] We recognize that the analysis of Dupas et al. (2021) spans many universities and the spatial component increased their required number of coders. Nevertheless, we argue that the algorithmic approach is still more efficient with human labor considering there are many NBER SI sessions that happen concurrently.

2 Data and Measurement

In this section we discuss our data sources for both training our neural network and for our analysis of economic presentations.

2.1 Data Sources

In this paper we use publicly available, labeled audio data to develop our machine learning algorithm and apply our algorithm to recordings of economic presentations from the NBER SI from the summer of 2022. To build our algorithm, we use two datasets commonly used in the computer science literature: the Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS) and the Crowd-sourced Emotional Multimodal Actors Dataset (CREMA-D) (Cao et al., 2014; Livingstone and Russo, 2018). Broadly speaking, these datasets include audio recordings of hired actors speaking in the tone they were instructed to use—happy, neutral, angry, sad, disgusted, or fearful—and information on the actors' gender and age. Across the datasets, we have about 8,500 short recordings. The data are balanced between male and female actors, and the ages range from 20 to 74. While there are some variations in accent, the datasets feature predominantly English words spoken in a neutral American accent.

For our actual analysis, we use recordings of presentations from the NBER SI, a major economics conference in the summer that spans three weeks with almost 500 presentations from a variety of fields listed in Table 1. One reason we study seminar dynamics at the NBER SI is that it allows us to have a representative sample of fields to study in our analysis of gender and tone in economics presentations. Another reason is that gender biases at high-stakes conferences have a clearer linkage to later-career outcomes. Presentations at the NBER SI are highly regarded because they are thought to have a strong positive effect on one's academic work. Session organizers are influential, and prominent economists in their fields and many top journal editors and referees participate as presenters, discussants, and audience members. Presenting at the NBER SI may affect your chances of publication, invitations to other seminars, and overall recognition in the profession. For example, Chari and Goldsmith-Pinkham (2017) find that papers in the NBER SI get published in top four journals at a higher rate.[10] Given the particular value our profession puts on presentations at the NBER SI, producing detailed measures of the dynamics in presentations is particularly important.

Due to the COVID-19 pandemic, the NBER SI was completely virtual for the 2020 and the 2021 conferences. Participants presented, discussed, and asked questions via Zoom. The 2022 NBER SI mostly returned to the in-person format but retained an option for participants to attend virtually over Zoom. The presentations were also broadcast live from the NBER's YouTube page, and the recordings remained available for two weeks. Before the conference, the NBER notified participants that the presentations would be available on their YouTube page and studied:

> *This year's meetings will be hybrid. The hybrid format offers expanded access for those who are not able to attend in person, but it also brings new challenges. Recordings of meetings will be posted on the NBER's YouTube channel for two weeks, where they can be accessed by students and other interested viewers. Audio recordings from the YouTube postings are also going to be transformed to anonymized data by a research team that is carrying out statistical analysis of conference dynamics.*

Even though these videos were temporarily publicly available online, there was an expectation of privacy in that the videos would eventually be taken down. Accordingly, we submitted our project to the Institutional Review Board (IRB) at both Brown University and NBER and received an exemption from review for our project due to the secondary nature of the data. We took many additional steps to protect the anonymity of participants at the NBER SI and to notify them that the recordings would be analyzed. The NBER SI organizers also notified participants that we would be analyzing presentation dynamics from

[10] Chari and Goldsmith-Pinkham (2017) find that 10 percent of papers at the NBER SI get into top four journals, while Card and DellaVigna (2013) find that top journals only accept 6 percent of submitted papers.

the recordings of presentations.[11]

The main unit of analysis for speaker dynamics is within a presentation. We do not track speakers across presentations, and within sessions, we use anonymized labels of "speaker 1" for the presenter, "speaker 2" for the discussant, and "speaker 3, i" for the ith utterance by an audience member. Furthermore, it is important to note that we can only create labels for audience members who spoke, and thus we cannot speak to the composition of audience members who were silent.

For presentation identifiers, we first aggregate sessions up to broad fields of "macroeconomics" and "microeconomics" and only keep the broader field labels for each presentation. We also document whether the presentation has a discussant format or a regular seminar-style format, as this logically has an effect on presentation dynamics. Unlike the previous literature, we do not have a separate "finance" field because there are not enough participants in each combination of field and presentation format, and thus we decrease the number of field categories.

Subsequently, once we impute the labels for tone, gender, and age from feeding the audio recordings into our neural network, we have our anonymized dataset. In the next section we talk more about this imputation process. Once we have our labels, we delete all the audio recordings and any linkages that can reconnect the data to the presentation session or speaker in accordance with our agreement with NBER. We take privacy concerns seriously in that we hope that our research will help support more equitable presentation experiences. In the next section, we detail our machine learning approach to imputing gender, tone, and age labels from raw audio recordings.

2.2 Imputation of Audio Labels

We impute gender, age, and tone labels for audio clips from the 2022 NBER SI given the speaker roles we have hand classified. We first use a software transcription service called

[11] Participants were given the option to not participate in our study. We were not given recordings for participants who opted out, and therefore they are not included in our analysis.

Trint to help us identify speakers and when speakers alternate. We then use convolutional neural networks to impute the gender, age, and tone for each speaker and their utterances. Specifically, we build gender-specific convolution neural networks to account for possible algorithmic gender biases toward physiologically different frequency ranges across genders. We go over the structure of our models and our imputation procedures in detail below.

To identify speakers and classify their roles, we use Trint (2022). Trint transcribes audio clips and tags the different speakers based on pauses and shifts in voice characteristics. Following the initial speaker tagging given by Trint, we manually go through all the audio recordings and identify speakers as presenters, discussants, or audience members. Although the transcript might offer useful information on how people interact, we ignore it for both privacy and accuracy concerns. The algorithm that Trint uses is not likely trained on economic speech and accordingly mistranslates many words.[12] With the speaker switches identified, we assign an anonymous label of "speaker 1" for the presenter, "speaker 2" for the discussant, and "speaker 3, i" for the ith utterance by an audience member. In addition to speaker roles, we have the timestamps on when speakers switch. This allow us to split seminar data into speaker utterances and, subsequently, five-second intervals of speech.

To impute gender, age, and tone, we train a series of convolutional neural networks that classify gender, age, and tone of speech. Convolutional neural networks are a class of artificial neural networks, and they capture hierarchical relationships by applying layers of convolutional kernels to a multi-dimensional feature matrix. The convolution operation multiplies the kernels with local regions of the feature matrix element by element and sums their products. This sum captures key information to summarize that local region. In our application to audio data, these local regions can represent shifts in intonation or pitches, whereas in images, the local regions can signify lines and edges. We consider this method to be a great alternative to the multi-layer perceptron approach taken in Gorodnichenko et al. (2021), as convolutions allow for local shift invariance, which means the model predictions

[12] For example, "shocks" are often mistranslated as "sharks," "MPK" is translated as "Olympic," and "Cournot" becomes "quinoa."

are robust to small shifts in input features.

The network structure we choose includes two layers of one-dimensional convolutional layers, followed by a pooling layer and two more one-dimensional convolutional layers. This hierarchy of convolutional layers can be seen as detecting feature characteristics at a gradually more general level, going from small local fluctuations in pitches to an overall classification of tones or age. We iterate the models over our training data 200 times, minimizing the cross-entropy between the predicted labels and the true labels for classification tasks (gender and tone) and the mean squared errors for regression tasks (age). We have three sets of models, trained to predict age, emotions, and gender. For age and emotion, we train two gender-specific models for each label. We do this because we want to eliminate potential gender biases, such as classifying audio clips with a happier pitch as angry or happy, while women speak at higher frequency domains than men. Thus, we have a total of five models: two for age, two for emotion, and one for gender.

To impute age and tones for each gender, we must first classify the speaker's gender. While an inaccurate gender classifier might compound errors in downstream analysis, luckily, gender classification is a relatively easy task, and we achieve near perfect accuracy in our validation data. We feed our five-second audio clips into the gender classifier and classify a speaker's gender by taking the mode of predictions for all the audio clips spoken by the speaker. Based on gender predictions, we then feed the audio clips into respective gender-specific models to impute the probability distribution over the emotions, as well as the age, which is continuous.

To ensure our predictions are robust, we cross-validate our model and split our datasets into five folds of training and validation data. We then train a separate set of models for each fold of the train/validation "splits." At inference, we take the average of the predictions generated by each of the five folds. To summarize, we first predict gender using the average prediction from all five folds of the train/validation splits. We then apply gender-specific models for tone and age for all five folds and average the prediction labels. Table 2 shows the accuracy and performance results averaged across all five folds of the models.

2.3 Additional Pre-Processing

In the 2022 NBER SI, there are three types of presentations: presentation with discussants, regular seminar format presentations, and lightning round presentations. Since our goal is to capture how gender impacts interactions in economics seminars, we drop all lightning round presentations. We also drop all presentations that have multiple presenters because there are too few presentations of this type that it could have been possible for someone to identify the presentation. We group the presentations into two fields: macroeconomics and microeconomics (henceforth, macro and micro). We choose not to include finance as a separate field due to identifiability concerns and instead manually group the finance sessions into micro and macro sessions. Table 1 documents how we group the programs. The resulting dataset includes a total of 479 presentations, 234 of which are micro presentations.[13]

3 Empirical Analysis

Our empirical analysis has five steps. We first look at the representation of male and female economists as presenters, discussants, and audience members based on our imputed gender variable. Then, we examine the cross-section of tones across speakers and analyze how speaker characteristics and the field and seminar format impact how happy or sad the speakers sound. We next look at the within-speaker dynamic and document the transition and persistence of tones, and then look at the number of interruptions in a presentation. Last, we examine the between-speaker interaction dynamics by looking at the impact of the previous speaker's tone on the current speaker and how this impact systematically differs across speaker types, fields, and seminar types.

[13] Sessions that are on the 2022 NBER SI program but are not on our list included participants who opted out of having their session recorded.

We conduct our analysis at three levels of aggregations, which stem from our data-processing procedures. Through our processing, we split presentations into uninterrupted speech segments, which we call "utterances." Utterances are uniquely identified by the order in which they appear in the presentations, and each one is assigned a speaker ID. They vary in time—an utterance can be a 20-minute uninterrupted presentation or a "thank you" statement that lasts for a few seconds. Since our models are trained on audio clips three to five seconds long, we match the time duration during our imputation by splitting all utterances into five-second splits. We generate labels for speaker- and utterance-level observations by taking the average of the labels imputed at the five-second level across all splits within the utterance or spoken by the speaker. Each level of aggregation comes with its advantages and disadvantages, and we dive into the details when we go over the regressions.

3.1 Gender Representation

Overall, we find that about 60 percent of presenters are male and 40 percent are female. When we break this down by field, the gender split in macro is almost equal—55 percent male and 45 percent female— while the split in micro favors men—66 percent male and 33 percent female. Overall, the NBER speaker female share is larger than the female share by rank for tenure-track positions documented in Chari and Goldsmith-Pinkham (2017), which is below 30 percent for all economics departments and is even lower for top departments. This finding suggests the program organizer put in effort to promote inclusion and equal representation in the economics profession. Table 3 summarizes the distribution of gender at the NBER SI.

Despite the increased number of female presenters, we still see segmentation of discussants and audience members. Figure 1 documents the gender share of participants by presenter gender and field. When the presenter is male, 70–80 percent of the discussants and audience members who speak are men. In contrast, there is a higher fraction of female discussants and female audience members who speak when there is a female presenter.

When we remove presentations where only one gender speaks, we find that the share of female-speaking audience members remains the same for macro talks but the gap for micro talks disappears. This indicates that there is likely more unequal selection into micro sub-fields by gender than in macro. Due to privacy concerns, we do not keep track of sub-fields or individual session effects. To proxy for female-dominated sub-fields, we look at the distribution of the female share of speakers over all of the NBER SI. We then flag presentations with female shares over the 75th percentile as belonging to a female-dominant talk.[14] Table 4 shows the regression results with speaking audience members' gender in macro and micro talks as the dependent variable. Even when controlling for the female-dominant talk, we still find an increased probability of a speaking audience member being female if there is a female presenter.

Furthermore, the share of audience members and discussants of the presenter's own gender is still higher for male presenters than for female presenters for both fields. This likely reflects that there are still more men in the economics profession than women. It is worth noting that our approach relies on measuring audience participation by those who speak. It is possible that presentation attendance with those who did not speak could be more balanced or imbalanced than what we find.

3.2 Cross-Sectional Tone Analysis

Our measure of tone is imputed from the raw audio data and our trained neural network algorithm. For each five-second interval, we have a probability distribution over six tones: sad, angry, neutral, happy, disgusted, and fearful. The summary statistics of these probabilities are described in Table 5. We view these as likelihoods of tone. Alternatively, one can consider these probabilities as the intensity measures of each tone. We use both interpretations in our discussion interchangeably.

We analyze how the cross-section of tones depends on speaker characteristics and field

[14] The 75th percentile is a presentation where over 66 percent of the speakers are female. In later tone regressions, we add female share as a control.

and seminar types. Our underlying relationship of interest is

$$\mathrm{T}_t^i = \beta_f f^i + \beta_a a^i + \beta_m \mathrm{m}^i + \beta_r r^i + \beta_s s \epsilon_t^i, \tag{1}$$

where T denotes tone, f the female indicator, a the age, m the macro field indicator, s the share of female speakers in the presentation, and r the regular seminar format indicator. Note that we include the female share to control for the possible selection effects that certain sub-fields have more women in general. Additionally, i denotes speaker-level characteristics, and t denotes observation at the five-second level. We assume ϵ_{it} to be i.i.d with mean zero.

Our machine learning models provide unbiased estimates $\hat{f^i}, \hat{T_t^i}$, and $\hat{a}^i$, where

$$\begin{aligned}
\hat{f^i} &= f^i + \epsilon_{f_i}, \\
\hat{a}^i &= a^i + \epsilon_{fa_i}, \\
\hat{T_t^i} &= T_t^i + \epsilon_{fT_{it}}.
\end{aligned}$$

We assume $\mathbb{E}[\epsilon_{f_i}] = \mathbb{E}[\epsilon_{fa_i}] = \mathbb{E}[\epsilon_{fT_{it}}] = 0$. Note that we include f, which is the female indicator in the error terms for $\hat{a}^i$ and $\hat{T_t^i}$, since the gender-specific models we use to predict tone and age depend on predictions from the gender classifier. We can rewrite our relationship of interest as

$$\hat{T_t^i} = \beta_f \hat{f^i} + \beta_a \hat{a}^i + \beta_m m^i + \beta_r r^i - \beta_f \epsilon_{f_i} - \beta_a \epsilon_{fa_i} + \epsilon_{fT_{it}} + \epsilon_t^i. \tag{2}$$

Here, a natural concern is that the error terms in $\hat{a}^i$ and $\hat{T_t^i}$ are correlated with ϵ_{fi}. However, as mentioned earlier, the gender classification model has a near 100 percent accuracy, and we assume that $\hat{f^i} = f^i$, meaning that $\epsilon_{fi} = 0$ is uncorrelated with ϵ_{fai} and ϵ_{fTit}. Additionally, we assume ϵ_{fai} and ϵ_{fTit} are independent, which guarantees consistent and unbiased estimates of the βs. Note that we can generalize this to the utterance level and the speaker level, as utterance- and speaker-level tone labels are averages of the five-second-level tone labels.

Table 6 documents the estimation results for Equation 2 that illustrate how presenter tones depend on presenter characteristics at the speaker level. It shows that female presenters are on average 11 percentage points more likely to sound happy. Interestingly, presenters who sound older are 2 percentage points less likely to sound happy. Note that we standardize ages by gender, so the coefficients capture the correlation between a speaker's tones and how many standard deviations away they sound compared to the mean of their gender. In addition, presenters are 7.8 percentage points more likely to sound angry in regular seminars and are 7.5 percentage points less likely to sound angry in macro seminars. This set of findings reflects systematic differences in presentation styles across gender, age, field, and seminar formats. For example, female presenters need to sound happier when communicating their research, and presenters sound more serious when they are in regular seminars, which are usually associated with more questions and back-and-forths.

Table 7 records the impact of audience characteristics on audience tones. Similar to speaker tones, female audience members are 10.9 percentage points more likely to sound happy. This indicates that female audience members can be more supportive and encouraging to the presenters. Age now has an opposite effect on how likely the speaker is to sound happy: economists that are one standard deviation older than the mean age for their gender are 0.9 percentage points more likely to sound happy. Audience members in macro seminars are 5.4 percentage points less likely to sound happy than audience members in micro seminars. Interestingly, the share of female speakers in the presentations has a negative coefficient on *Happy*, suggesting a strange phenomenon that in environments where more speakers are female, audience members are less likely to sound happy.

Table 8 looks at how split-level cross-sectional tones depend on speaker, presentation characteristics, and tones at the last split. We find that our previous results at the speaker level remain qualitatively robust, as *Female* still remains a statistically predictor for happy tones. However, we also find that tones are persistent and explain a nontrivial fraction of the variability we observe in the data. We formally investigate how tones transition in the next section.

3.3 Within-Speaker Tone Dynamics

In examining the within-speaker dynamics of tones, we are interested in learning about how persistent the tones are and how they transition. We estimate a tone transition matrix for speakers of different genders by regressing current tone as a function of previous tones by estimating the following relationship:

$$T_t^{fi} = \boldsymbol{\beta}_{T_{\text{trans}}} \boldsymbol{T}_{t-1}^{fi} + \epsilon_t^i, \quad \text{for } f = 0 \text{ and } 1,$$

which can be rewritten as

$$\hat{T}_t^{fi} = \boldsymbol{\beta}_{T_{\text{trans}}} \hat{\boldsymbol{T}}_{t-1}^{fi} + \epsilon_t^i + \epsilon_{fTit} - \beta_{T_{\text{trans}}} \epsilon_{fTi,t-1}, \quad \text{for } f = 0 \text{ and } 1, \tag{3}$$

where $\boldsymbol{T}$ denotes the vector of probability distribution of tones and $\boldsymbol{\beta}_{T_{\text{trans}}}$ the vector of corresponding transition probabilities. In practice, we drop one tone to avoid perfect multicolinearity since the distribution of tones sums up to one.

Table 9 and Table 10 document the gender-specific emotion transition matrix by estimating Equation 3 with OLS. Angry, sad, and fearful seem to be more persistent for male speakers and happy and neutral for female speakers. Additionally, female speakers have a higher probability of switching from angry to happy, which suggests that they adjust more positively from angry tones. As we can see, lagged tones are powerful tone predictors, achieving R^2s ranging from 0.3 to 0.8.

3.4 Audience Interruptions

In this subsection we discuss patterns of interruptions to quantify how audiences interact with male versus female presenters. We examine three aspects of interruptions: the timing, the duration, and the number of interruptions. Here, interruptions can be either prompted or unprompted and thus cover the general category of audience comments and questions. To focus on interruptions directed toward the presenter, we do not count any audience interruptions that occur during discussions.

When we look at when interruptions happen, we find that the difference in timing is larger across seminar formats than across fields and genders. That is, we see that interruptions are more likely to occur near the end for presentations with discussants compared to more evenly throughout a presentation in a regular seminar format.

The first two columns of Table 11 formalize this idea by showing that the first position of interruptions are much earlier for macro seminars with regular format, and the average position of interruptions are earlier for regular seminars and even earlier for regular macro seminars. Although the regressions suggest that interruptions tend to occur later in sessions with female presenters, this effect is not statistically significant.

The third column of Table 11 examines how the time share of audience utterances depend on field, format, and speaker gender. We find that, on average, audience utterances account for 22 percent of the total duration of utterances by presenters and all audience members. This ratio is lower for presentations with discussants since sessions with discussions usually have a shorter presentation by the presenter. Here, we see that female presenters are likely to encounter a larger share of audience interruption. The coefficient for *Female* is significant at the 10 percent level.

We now turn to the number of interruptions. For discussion seminars, we count the total number of interjections before the discussants speak. We also count the total number of questions and comments given by audience members for both the discussion and the regular seminars during the entire presentation. Table 12 shows the regression results for the two measures of interruptions for both seminar formats. The first column shows there are, on average, around two interruptions for micro seminars with discussants and four for macro seminars before the discussant speaks. The second column shows there are an average of around 11 total interruptions for seminars with discussants, but the number of total interruption does not vary across fields and the presenter's gender. The third column shows that, for macro seminars with a regular format, there are, on average, 8 more questions or comments. Interestingly, none of the coefficients on $Female$ are statistically significant, suggesting that field and seminar formats have more impact on the number of questions or comments compared to the presenter's gender.

3.5 Between-Speaker Tone Dynamics

We now turn to how speakers of varying characteristics interact. In this section, we analyze how different type of speakers interact with each other and how these interactions depend on speaker types, genders, and seminar types. In addition, we identify audience interruptions that shift presenter tones, and we characterize the conditions under which more positive and negative tone shifts are present.

We are interested in the different interactions speakers of different roles have across different seminar types. In particular, for speaker i and j, where $i, j \in \{$Presenter, Discussant, Audience$\}$, we estimate the following:

$$\hat{T}^i_t = \beta_f \hat{f}^i + \beta_a \hat{a}^i + \beta_{f_{-1}} \hat{f}^j + \beta_{a_{-1}} \hat{a}^j + \beta_m m_i + \hat{\beta}_{\boldsymbol{T}_{-1}} \hat{\boldsymbol{T}}^j_{t-1} + \eta^i_t. \tag{4}$$

In words, we are interested in estimating how the tone of speaker i at time t depends on speaker i's characteristics, session characteristics, the previous speaker j's characteristics, and the tones of speaker j at time $t-1$.

Table 13 and Table 14 estimate a specific case of Equation 4 by looking at how speaker happy tones depend on characteristics of the current and the previous speakers for different seminar types. In each table, each column examines a particular type of interaction $i - j$, in which i is the role of the current period speaker and j is the role of the previous period speaker. For example, the column "Presenter-Audience" in Table 13 is the same as estimating

Equation 4 by setting $i =$ Presenter and $j =$ Audience. Additionally, we only consider cases where t is the first five seconds of speaker i's utterance and $t-1$ is the last five seconds of speaker j's utterance. In other words, we are only focusing on the short ten-second windows in which speakers switch. Our results are robust to changing the time window of consideration to the entire utterance.

Across specifications, female presenters and audiences are 10 to 47 percentage points more likely to sound happy, which reaffirms our previous results in Table 6 and Table 7. We also see a consistently negative coefficient for $Female\_prev$, which indicates that the current speaker is less likely to respond happily when the previous speaker is female. This effect is statistically significant and quantitatively large, suggesting there could be certain biases when responses toward the previous speaker could be systematically different and dependent on their gender. Somewhat surprisingly, in the "presenter-audience" regression results in Table 13 and Table 14, the coefficient for the interaction between $Female$ and $Female\_prev$ is negative, significant, and sizable, suggesting that female presenters are 7 to 12 percentage points less likely to respond in a happy tone to female audience members. One possible explanation is that female audience members are more critical toward female presenters.

Furthermore, we want to investigate under what circumstances the presenters experience large swings in their tones. To do this, we identify extreme tone shifts by looking at consecutive utterances following a "presenter-audience-presenter" pattern. This enables us to compute the change in tone for the presenter, which could be due to either random fluctuations or audience interruptions. We then take the top and the bottom deciles of the tone shifts, ranked by $\Delta\hat{T}^i = \hat{T}^i_{t+1} - \hat{T}^i_{t-1}$, where $i =$ presenter, across all presentations. We count the number of top positive and negative tone shifts that each presenter encounters. We then relate the number of extreme tone shifts to speaker and session characteristics, through the following regression:

$$\text{Count of Extreme Tone Shifts} = \beta_f \hat{f}^i + \beta_a \hat{a}^i + \beta_E E^i + \beta_s \hat{s}^i + \eta, \tag{5}$$

where i = presenter and s denotes the female share in the presentation that presenter i presents in. Specifically, we consider extreme positive tone shifts and negative ton shifts separately.

Table 15 estimates Equation 5 by looking at changes in the the probability of sounding happy. We want to focus on the shift in presenter tones that can possibly be attributed to audience member interruptions due to the timing of the event window. Thus, we compute $\Delta Happy^i$ by subtracting the $Happy$ probability at the last minute of a presenter's first utterance by their $Happy$ probability at the first minute of their second utterance. Otherwise, since presenter utterances can be long, especially if they are in the middle of giving a presentation, computing $\Delta Happy^i$ using whole utterances might dampen the effects the audience interruption has, although, as we show in the appendix, our results are robust to using whole utterances instead.

We find that presenters in macro are likely to experience more large positive and negative swings in how likely they are to sound happy, with a slightly larger coefficient for the number of most negative tone shifts. Additionally, female presenters tend to experience more large positive shifts in the likelihood of their happy tone. The reason for this might be either that female presenters are greeted with more encouragement or that they tend to respond more happily when interacting with audience members.

We also investigate how the fraction of interruptions that dramatically shift the presenter's tone depends on speaker and session characteristics. The results are summarized in Table 16. This controls for the fact that macro seminars are slightly longer and are more likely to have more interruptions for the same time period. Nevertheless, we still find macro presentations have more interruptions that cause the presenter to increase or decrease their probability of sounding happy drastically. That is, a macro talk relative to a micro talk will have 1.3 percentage points more extreme positive shifts and 2.9 percentage points more extreme negative shifts. However, when looking at the fraction of interruptions that are extreme tone shifts, we now find female presenters being more likely to have extreme negative tone shifts in addition to extreme positives shifts.

4 Discussion on Audio Analysis in Economics

In this section we discuss our recommendations for other researchers looking to perform automated audio analysis. In our paper, one of the key methodological approaches is to use gender-specific neural networks. Gender bias has been widely observed across areas of machine learning. For example, in generating word embedding, which is a popular method of representing text as numerical vectors, algorithms often learn to pick up gender stereotypes, such as associating homemakers with women (Bolukbasi et al., 2016). In speech emotion recognition, models often predict male emotions with higher accuracy, which can be due to imbalances in the training data as well as possible systematic differences in male and female voices (Gorrostieta et al., 2019).

We suggest that using gender-specific neural networks is a suitable way of mitigating such gender bias. To illustrate how male and female voices systematically differ, Figure 2 shows the confusion matrix of utterances by female speakers in CREMA-D and RAVDESS on a model that is trained on male-only utterances in the same dataset. The figure shows that, even controlling for the same data quality, models trained on male-only data exhibit systematic misclassification patterns when applied to female audio clips. Specifically, the male-only model overwhelmingly predicts "angry," "happy," and "disgust" for female speakers. Such differences can be due to difference in pitches and amplitudes in male- and female-speaking patterns. For comparison, Figure 3 shows the confusion matrix for a model trained on female-only data, evaluated on the validation set, which the model has never seen before.

Here, we see no obvious systematic prediction errors.

5 Conclusion

In this paper, we study the dynamics of tone in presentations at the 2022 NBER SI and how those dynamics interact with economist gender, economist age, presentation format, and the economic field. We use a scalable, time-consistent machine learning approach to label audio data with tone, age, and gender at five-second intervals. With this high-frequency measurement, we provide new results on the dynamics of tone in economics presentations. We find statistically significant differences in the tone both used by and directed toward female economists. We also find that female economists are more likely to sound positive and economists are more likely to sound negative responding to female economists. Moreover, we find that women also sound more negative when responding to women.

Going forward, we hope that others can use our trained models to assign comparable tone labels for future economics presentations, and, accordingly, we will make our algorithm publicly available. Alternatively, for economists who wish to develop their own audio-tone models, we outline best practices for researchers looking to impute labels from audio data on men and women. In short, we show that researchers should use gender-specific models and properly standardize their audio data to avoid biased labels.

Comparing our paper with the previous literature, we see there is improvement in terms of the representation of female economists and that there is more equality on some dimensions of presentations, like the number of interruptions. Studying the dynamics in presentations is only one part of the story, but it likely has spillovers into other dimensions of academic careers. We leave this connection of presentation quality to other academic outcomes to future research.

References

Abrevaya, Jason and Daniel S. Hamermesh (2012) "Charity and Favoritism in the Field: Are Female Economists Nicer (To Each Other)?" *Review of Economics and Statistics*, 94 (1), 202–207, 10.1162/rest_a_00163.

Alexopoulos, Michelle, Xinfen Han, Oleksiy Kryvtsov, and Xu Zhang (2022) "More Than Words: Fed Chairs' Communication During Congressional Testimonies," *Bank of Canada Working Paper Series*, 10.34989/swp-2022-20.

Alves, Jairo L., Leila Weitzel, Paulo Quaresma, Carlos E. Cardoso, and Luan Cunha (2019) "Brazilian Presidential Elections in the Era of Misinformation: A Machine Learning Approach to Analyse Fake News," in Nyström, Ingela, Yanio Hernández Heredia, and Vladimir Milián Núñez eds. *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, Lecture Notes in Computer Science, 72–84, Cham: Springer International Publishing, 10.1007/978-3-030-33904-3_7.

Avilova, Tatyana and Claudia Goldin (2018) "What Can UWE Do for Economics?" *AEA Papers and Proceedings*, 108, 186–90, 10.1257/pandp.20181103.

Aykanat, Murat, Özkan Kılıç, Bahar Kurt, and Sevgi Saryal (2017) "Classification of lung sounds using convolutional neural networks," *EURASIP Journal on Image and Video Processing*, 2017 (1), 65, 10.1186/s13640-017-0213-2.

Badshah, Abdul Malik, Jamil Ahmad, Nasir Rahim, and Sung Wook Baik (2017) "Speech Emotion Recognition from Spectrograms with Deep Convolutional Neural Network," in *2017 International Conference on Platform Technology and Service (PlatCon)*, 1–5, February, 10.1109/PlatCon.2017.7883728.

Bisbee, James, Nicolo Fraccaroli, and Andreas Kern (2022) "Yellin' at Yellen: Gender Bias in the Federal Reserve Congressional Hearings," February, 10.2139/ssrn.4030121.

Bolukbasi, Tolga, Kai-Wei Chang, James Zou, Venkatesh Saligrama, and Adam Kalai (2016) "Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings," *Working Paper*, 10.48550/ARXIV.1607.06520.

Bosquet, Clément, Pierre-Philippe Combes, and Cecilia García-Peñalosa (2019) "Gender and Promotions: Evidence from Academic Economists in France*," *The Scandinavian Journal of Economics*, 121 (3), 1020–1053, 10.1111/sjoe.12300.

Camastra, Francesco and Alessandro Vinciarelli (2015) *Machine Learning for Audio, Image and Video Analysis: Theory and Applications*: Springer.

Cao, Houwei, David G. Cooper, Michael K. Keutmann, Ruben C. Gur, Ani Nenkova, and Ragini Verma (2014) "CREMA-D: Crowd-sourced Emotional Multimodal Actors Dataset," *IEEE Transactions on Affective Computing*, 5 (4), 377–390, 10.1109/TAFFC.2014.2336244.

Card, David and Stefano DellaVigna (2013) "Nine Facts about Top Journals in Economics," *Journal of Economic Literature*, 51 (1), 144–161, 10.1257/jel.51.1.144.

Card, David, Stefano DellaVigna, Patricia Funk, and Nagore Iriberri (2020) "Gender Differences in Peer Recognition by Economists," *Quarterly Journal of Economics*, 135, 269–327.

——— (2022) "Gender Differences in Peer Recognition by Economists," *Econometrica*, 90, 1937–1971.

Chari, Anusha and Paul Goldsmith-Pinkham (2017) "Gender Representation in Economics Across Topics and Time: Evidence from the NBER Summer Institute," *NBER Working Paper Series* (23953), 10.3386/w23953.

Chevalier, Judy, Alicia Modestino, Pascaline Dupas, Jennifer Doleac, and Heather Tookes (2021) "Disparities in Economics Seminars: Research and Responses," https://www.aeaweb.org/about-aea/committees/cswep/programs/resources/webinars/disparities-2021.

CSWEP Committee on the Status of Women in the Economics Profession (2022) "CSWEP News," www.aeaweb.org/about-aea/committees/cswep.

Dolado, Juan J., Florentino Felgueroso, and Miguel Almunia (2011) "Are men and women-economists evenly distributed across research fields? Some new empirical evidence," *SERIEs*, 3 (3), 367–393, 10.1007/s13209-011-0065-4.

Doleac, Jennifer L., Erin Hengel, and Elizabeth Pancotti (2021) "Diversity in Economics Seminars: Who Gives Invited Talks?" *AEA Papers and Proceedings*, 111, 55–59, 10.1257/pandp.20211084.

Dupas, Pascaline, Alicia Modestino, Muriel Niederle, Justin Wolfers, and The Seminar Dynamics Collective (2021) "Gender and the Dynamics of Economics Seminars," *NBER Working Paper Series* (28494), 10.3386/w28494.

Ginther, Donna K. and Shulamit Kahn (2021) "Women in Academic Economics: Have We Made Progress?" *AEA Papers and Proceedings*, 111, 138–142, 10.1257/pandp.20211027.

Gorodnichenko, Yuriy, Tho Pham, and Oleksandr Talavera (2021) "The Voice of Monetary Policy," *NBER Working Paper Series* (28592), 10.3386/w28592.

Gorrostieta, Cristina, Reza Lotfian, Kye Taylor, Richard Brutti, and John Kane (2019) "Gender De-Biasing in Speech Emotion Recognition," in *Proc. Interspeech 2019*, 2823–2827, 10.21437/Interspeech.2019-1708.

Hastie, Trevor, Robert Tibshirani, and J. H. Friedman (2009) *The elements of statistical learning: data mining, inference, and prediction*, Springer series in statistics, New York, NY: Springer, 2nd edition.

Hengel, Erin (2022) "Publishing While Female: are Women Held to Higher Standards? Evidence from Peer Review," *The Economic Journal*, 132 (648), 2951–2991, 10.1093/ej/ueac032.

Hengel, Erin and Eunyoung Moon (2022) "Gender and equality at top economics journals," *Working Paper*, https://erinhengel.github.io/Gender-Quality/quality.pdf.

Hengel, Erin and Sara Louisa Phythian-Adams (2022) "An historical portrait of female economists' co-authorship networks," *History of Political Economy*, https://erinhengel.github.io/hope/hope.pdf.

Issa, Dias, M. Fatih Demirci, and Adnan Yazici (2020) "Speech emotion recognition with deep convolutional neural networks," *Biomedical Signal Processing and Control*, 59, 101894, 10.1016/j.bspc.2020.101894.

James, Anusha, Mohan Kashyap, Yi Han Victoria Chua, Tomasz Maszczyk, Ana Moreno Núñez, Rebecca Bull, and Justin Dauwels (2018) "Inferring the Climate in Classrooms from Audio and Video Recordings: A Machine Learning Approach," in *2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*, 983–988, December, 10.1109/TALE.2018.8615327, ISSN: 2470-6698.

Kahl, Stefan, Amanda Navine, Holger Klinck et al. (2022) "Overview of BirdCLEF 2022: Endangered bird species recognition in soundscape recordings," 08.

Liem, Cynthia C. S., Markus Langer, Andrew Demetriou, Annemarie M. F. Hiemstra, Achmadnoer Sukma Wicaksana, Marise Ph. Born, and Cornelius J. König (2018) "Psychology Meets Machine Learning: Interdisciplinary Perspectives on Algorithmic Job Candidate Screening," in Escalante, Hugo Jair, Sergio Escalera, Isabelle Guyon, Xavier Baró, Yağmur Güçlütürk, Umut Güçlü, and Marcel van Gerven eds. *Explainable and Interpretable Models in Computer Vision and Machine Learning*, 197–253: Springer International Publishing, 10.1007/978-3-319-98131-4_9.

Lim, Wootaek, Daeyoung Jang, and Taejin Lee (2016) "Speech emotion recognition using convolutional and Recurrent Neural Networks," in *2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, 1–4, December, 10.1109/APSIPA.2016.7820699.

Livingstone, Steven R. and Frank A. Russo (2018) "Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English," *PLOS ONE*, 13 (5), e0196391, 10.1371/journal.pone.0196391.

Lundberg, Shelly (2022) "Gender Economics and the Meaning of Discrimination," *AEA Papers and Proceedings*, 112, 588–591, 10.1257/pandp.20221086.

Naim, Iftekhar, M. Iftekhar Tanveer, Daniel Gildea, and Mohammed Ehsan Hoque (2015) "Automated prediction and analysis of job interview performance: The role of what you say and how you say it," in *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, 1, 1–6, May, 10.1109/FG.2015.7163127.

Nguyen, Laurent Son, Denise Frauendorfer, Marianne Schmid Mast, and Daniel Gatica-Perez (2014) "Hire me: Computational Inference of Hirability in Employment Interviews Based on Nonverbal Behavior," in *IEEE Transactions on Multimedia*, 16, 1018–1031, June, 10.1109/TMM.2014.2307169.

Sarsons, Heather (2017) "Recognition for Group Work: Gender Differences in Academia," *American Economic Review*, 107 (5), 141–145, 10.1257/aer.p20171126.

Sarsons, Heather, Klarita Gërxhani, Ernesto Reuben, and Arthur Schram (2021) "Gender Differences in Recognition for Group Work," *Journal of Political Economy*, 129 (1), 101–147, 10.1086/711401.

Sarsons, Heather and Guo Xu (2021) "Confidence Men? Evidence on Confidence and Gender among Top Economists," *AEA Papers and Proceedings*, 111, 65–68, 10.1257/pandp.20211086.

Sierminska, Eva and Ronald L. Oaxaca (2021) "Field Specializations among Beginning Economists: Are There Gender Differences?" *AEA Papers and Proceedings*, 111, 86–91, 10.1257/pandp.20211030.

Story, Brad H. and Kate Bunton (2015) "Formant measurement in children's speech based on spectral filtering," *Speech Communication*, 76, 93–111, 10.1016/j.specom.2015.11.001.

Teodorescu, Mike, Nailya Ordabayeva, Marios Kokkodis, Abhishek Unnam, and Varun Aggarwal (2022) "Determining systematic differences in human graders for machine learning-based automated hiring," *Brookings Working Paper Series*.

Trint (2022) "Transcribe video and audio to text with AI," https://trint.com.

Vogt, Thurid and Elisabeth André (2006) "Improving Automatic Emotion Recognition from Speech via Gender Differentiaion," in *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, Genoa, Italy: European Language Resources Association (ELRA), May, http://www.lrec-conf.org/proceedings/lrec2006/pdf/392_pdf.pdf.

Xia, Tong, Jing Han, and Cecilia Mascolo (2022) "Exploring machine learning for audio-based respiratory condition screening: A concise review of databases, methods, and open issues," *Experimental Biology and Medicine*, 15353702221115428, 10.1177/15353702221115428.

Zeng, Yuni, Hua Mao, Dezhong Peng, and Zhang Yi (2019) "Spectrogram based multi-task audio classification," *Multimedia Tools and Applications*, 78 (3), 3705–3722, 10.1007/s11042-017-5539-3.

Zhang, Shiqing, Shiliang Zhang, Tiejun Huang, and Wen Gao (2018) "Speech Emotion Recognition Using Deep Convolutional Neural Network and Discriminant Temporal Pyramid Matching," in *IEEE Transactions on Multimedia*, 20, 1576–1590, June, 10.1109/TMM.2017.2766843.

Zisad, Sharif Noor, Mohammad Shahadat Hossain, and Karl Andersson (2020) “Speech Emotion Recognition in Neurological Disorders Using Convolutional Neural Network,” in Mahmud, Mufti, Stefano Vassanelli, M. Shamim Kaiser, and Ning Zhong eds. *Brain Informatics*, Lecture Notes in Computer Science, 287–296, Cham: Springer International Publishing, 10.1007/978-3-030-59277-6_26.

A Tables

Table 1: NBER Summer Institute Sessions by Field

Macro	**Micro**
Asset Pricing	Aging
Behavioral macro	Big Data and High-Performance Computing for Financial Economics
Capital Markets and the Economy	Children
Conference on Research in Income and Wealth	Development Economics
Corporate Finance	Economics of National Security
Development of the American Economy	Economics of Social Security
Dynamic Equilibrium Models	Entrepreneurship
Economic Fluctuations and Growth	Environmental & Energy Economics
Economic Growth	Gender in the Economy: Change and Persistence of Norms
Forecasting & Empirical Methods	Health Care
Impulse and Propagation Mechanisms	Health Economics
Income Distribution and Macroeconomics	Household Finance
International Finance & Macroeconomics	Industrial Organization
International Finance and Macroeconomic Data Sources	Innovation
International Trade & Macroeconomics	IT and Digitization
International Trade and Investment	Labor Studies
Macro Perspectives	Law and Economics
Macro, Money and Financial Frictions	Personnel Economics
Macroeconomics and Productivity	Political Economy
Macroeconomics Within and Across Borders	Public Economics
Micro Data and Macro Models	Real Estate
Monetary Economics	Science of Science Funding
Risk of Financial Institutions	Urban Economics

Note: the breakdown of these sessions largely follows Chari and Goldsmith-Pinkham (2017). We do not include a separate "Finance" group due to identifiability concerns. We reallocate the finance sessions into micro and macro according to the focus of the session. Sessions that are on the 2022 NBER SI program but are not on our list included participants that opted out of having their session recorded.

Table 2: Model Accuracy on Validation Set

	Emotion	Age	Gender
Both	-	-	96.47%
			(0.007)
Male	48.63%	94.65	-
	(0.019)	(14.016)	
Female	52.74%	99.29	-
	(0.028)	(15.285)	

Note: The performance for emotion and gender is evaluated based on accuracy, and that for age is evaluated based on root-mean-square error. The standard deviations are in the parentheses. Note that there are six available emotions, so our models significantly outperform random selection, which would give us accuracy around 16.67%.

Table 3: Summary of Speakers by Role, Format, and Field

		Macro		
Format	Gender	Presenter	Discussant	Audience
Discussion	Male	68	67	857
	Female	49	50	531
Regular	Male	66	-	1345
	Female	62	-	1238

		Micro		
Format	Gender	Presenter	Discussant	Audience
Discussion	Male	94	91	1006
	Female	41	44	471
Regular	Male	60	-	714
	Female	39	-	501

Note: Gender is imputed from audio data and our algorithm.

Table 4: Female Probability of Audience Participation

	Audience Participant, Female	
	Macro	Micro
Intercept	0.210***	0.214***
	(0.013)	(0.012)
Female Presenter	0.199***	0.068***
	(0.019)	(0.020)
Female-Dominant Talk	0.408***	0.611***
	(0.020)	(0.023)
Regular	0.014	-0.022
	(0.014)	(0.016)
Presenter Age	0.006	0.010
	(0.011)	(0.013)
R^2	0.301	0.324
Adj. R^2	0.300	0.323
N	3971	2692

Note: The gender and age variables are imputed using our algorithms. The dependent variable is a dummy variable the is one if the speaking audience member is female. The female dominant session dummy is one if the female share of speakers in the presentation is over 75th percentile of all presentations. We use this as a proxy for talks in female dominated sub-fields. The unit of analysis here is at the audience interruption level. * denotes $p < 0.05$, ** denotes $p < 0.01$, and *** denotes $p < 0.001$. The standard errors are OLS standard errors.

Table 5: Tone Summary Statistics

	Happy	Angry	Sad	Neutral	Fearful
Count	7395	7395	7395	7395	7395
Mean	0.298	0.287	0.097	0.080	0.094
Stdv	0.210	0.253	0.142	0.106	0.121
Min	0.000	0.000	0.000	0.000	0.000
Median	0.262	0.228	0.033	0.032	0.050
Max	0.995	1.000	0.951	0.644	0.985

Note: Tone is imputed using our machine learning algorithm.

Table 6: Probability of Presenter Tone

	happy	angry	sad	neutral	fearful
Intercept	0.3191***	0.3966***	0.0330***	0.0146***	0.1449***
	(0.0174)	(0.0209)	(0.0095)	(0.0052)	(0.0082)
Female	0.1124***	0.0242	-0.0496***	0.0018	-0.0970***
	(0.0229)	(0.0275)	(0.0125)	(0.0068)	(0.0108)
Age	-0.0214*	-0.0308**	0.0099	0.0079**	0.0155***
	(0.0122)	(0.0146)	(0.0066)	(0.0036)	(0.0057)
Macro	0.0179	-0.0749***	0.0308***	0.0272***	-0.0146
	(0.0217)	(0.0260)	(0.0118)	(0.0064)	(0.0102)
Regular	-0.0334	0.0778***	-0.0087	-0.0058	-0.0207*
	(0.0227)	(0.0273)	(0.0124)	(0.0067)	(0.0107)
Macro:Regular	-0.0199	-0.0385	0.0040	0.0167*	-0.0005
	(0.0314)	(0.0377)	(0.0171)	(0.0093)	(0.0148)
Female Share	-0.0252	-0.1692***	0.0935***	0.0383***	-0.0016
	(0.0363)	(0.0437)	(0.0198)	(0.0108)	(0.0171)
R^2	0.09	0.13	0.09	0.20	0.28
Adj. R^2	0.07	0.12	0.07	0.19	0.27
N	479	479	479	479	479

Note: This table is created by looking at the cross section of emotions for presenters only at the speaker level. The gender and age variables are imputed using our algorithms. * denotes $p < 0.05$, ** denotes $p < 0.01$, and *** denotes $p < 0.001$. The standard errors are OLS standard errors.

Table 7: Probability of Audience Tones

	happy	angry	sad	neutral	fearful
Intercept	0.2846***	0.4033***	0.0355***	0.0324***	0.1657***
	(0.0059)	(0.0068)	(0.0040)	(0.0029)	(0.0031)
Female	0.1090***	-0.1076***	0.0052	0.0511***	-0.0986***
	(0.0067)	(0.0077)	(0.0045)	(0.0033)	(0.0036)
Age	0.0090***	-0.0281***	-0.0020	-0.0062***	0.0070***
	(0.0025)	(0.0028)	(0.0017)	(0.0012)	(0.0013)
Macro	-0.0540***	-0.0756***	0.0747***	0.0543***	-0.0177***
	(0.0077)	(0.0089)	(0.0052)	(0.0038)	(0.0041)
Regular	-0.0038	0.0373***	-0.0079	-0.0196***	-0.0128***
	(0.0080)	(0.0092)	(0.0054)	(0.0039)	(0.0042)
Macro:Regular	0.0325***	0.0102	-0.0367***	-0.0089*	-0.0117**
	(0.0105)	(0.0121)	(0.0071)	(0.0052)	(0.0056)
Female Share	-0.0368***	-0.1354***	0.0904***	0.0294***	-0.0207***
	(0.0104)	(0.0120)	(0.0071)	(0.0051)	(0.0055)
R^2	0.06	0.15	0.09	0.14	0.22
Adj. R^2	0.06	0.15	0.09	0.14	0.22
N	6663	6663	6663	6663	6663

Note: This table is created by looking at the cross section of emotions for audience members only. The gender and age variables are imputed using our algorithms. * denotes $p < 0.05$, ** denotes $p < 0.01$, and *** denotes $p < 0.001$. The standard errors are OLS standard errors.

Table 8: Probability of Presenter Tones with Lagged Tone

	Happy_t	Angry_t	Sad_t	Neutral_t	Fearful_t
Intercept	0.1583***	0.1179***	0.0599***	0.0401***	0.0511***
	(0.0032)	(0.0032)	(0.0015)	(0.0012)	(0.0014)
Female	0.0334***	0.0109***	-0.0170***	0.0055***	-0.0385***
	(0.0016)	(0.0016)	(0.0008)	(0.0006)	(0.0007)
Age	-0.0051***	-0.0133***	0.0005	0.0024***	0.0070***
	(0.0008)	(0.0008)	(0.0004)	(0.0003)	(0.0004)
Macro	0.0024	-0.0173***	0.0071***	0.0097***	-0.0042***
	(0.0015)	(0.0015)	(0.0007)	(0.0006)	(0.0007)
Regular	-0.0131***	0.0300***	-0.0007	-0.0013**	-0.0108***
	(0.0015)	(0.0015)	(0.0007)	(0.0006)	(0.0007)
Macro:Regular	-0.0040**	-0.0184***	-0.0019**	0.0047***	0.0016*
	(0.0020)	(0.0020)	(0.0010)	(0.0007)	(0.0009)
Happy_{t-1}	0.5819***	0.0067*	-0.0426***	-0.0286***	0.0008
	(0.0035)	(0.0035)	(0.0017)	(0.0013)	(0.0016)
Angry_{t-1}	-0.0382***	0.6859***	-0.0695***	-0.0471***	0.0095***
	(0.0034)	(0.0034)	(0.0016)	(0.0013)	(0.0015)
Sad_{t-1}	-0.0675***	-0.1081***	0.6015***	0.0478***	-0.0114***
	(0.0051)	(0.0051)	(0.0025)	(0.0019)	(0.0023)
Neutral_{t-1}	0.0117	0.0271***	-0.0029	0.4501***	0.0171***
	(0.0071)	(0.0071)	(0.0034)	(0.0027)	(0.0033)
Fearful_{t-1}	-0.0496***	0.0630***	-0.0438***	-0.0377***	0.6186***
	(0.0048)	(0.0048)	(0.0023)	(0.0018)	(0.0022)
Female Share	-0.0087***	-0.0525***	0.0251***	0.0069***	0.0037***
	(0.0024)	(0.0024)	(0.0011)	(0.0009)	(0.0011)
R^2	0.40	0.53	0.49	0.37	0.47
Adj. R^2	0.40	0.53	0.49	0.37	0.47
N	179537	179537	179537	179537	179537

Note: This table is created by looking at the cross section of emotions for presenters only at the split level, controlling for within-speaker lagged tones from the previous split. The gender and age variables are imputed using our algorithms. * denotes $p < 0.05$, ** denotes $p < 0.01$, and *** denotes $p < 0.001$. The standard errors are OLS standard errors.

Table 9: Autocorrelation of Tone Probabilities for Male Speakers

	Happy_t	Angry_t	Sad_t	Neutral_t	Fearful_t
Happy_{t-1}	0.7614***	0.1224***	0.0345***	0.0318***	0.0518***
Angry_{t-1}	0.1055***	0.8178***	0.0013**	0.0047***	0.0534***
Sad_{t-1}	0.0784***	-0.0251***	0.7833***	0.0734***	0.0350***
Neutral_{t-1}	0.2680***	0.2529***	-0.0208***	0.5331***	0.0749***
Fearful_{t-1}	0.1090***	0.1442***	0.0228***	0.0104***	0.6852***
R^2	0.78	0.81	0.64	0.44	0.65
Adj. R^2	0.78	0.81	0.64	0.44	0.65
N	156388	156388	156388	156388	156388

Note: This table documents the tone transition for all male speakers at the split level. The gender and age variables are imputed using our algorithms. * denotes $p < 0.05$, ** denotes $p < 0.01$, and *** denotes $p < 0.001$. This relationship is estimated by OLS. 'disgut_prev' is not included as a regressor because of perfect multicolinearity.

Table 10: Autocorrelation of Tone Probabilities for Female Speakers

	Happy_t	Angry_t	Sad_t	Neutral_t	Fearful_t
Happy_{t-1}	0.7757***	0.1107***	0.0268***	0.0323***	0.0174***
Angry_{t-1}	0.1923***	0.7774***	-0.0006	0.0094***	0.0200***
Sad_{t-1}	0.1187***	-0.0146***	0.5805***	0.1767***	0.0397***
Neutral_{t-1}	0.2119***	0.0651***	0.2188***	0.5845***	0.0223***
Fearful_{t-1}	0.1524***	0.3298***	0.1039***	-0.0009	0.3625***
R^2	0.75	0.72	0.53	0.56	0.26
Adj. R^2	0.75	0.72	0.53	0.56	0.26
N	100970	100970	100970	100970	100970

Note: This table documents the tone transition for all female speakers at the split level. The gender and age variables are imputed using our algorithms. * denotes $p < 0.05$, ** denotes $p < 0.01$, and *** denotes $p < 0.001$. This relationship is estimated by OLS. 'disgut_prev' is not included as a regressor because of perfect multicolinearity.

Table 11: Interruption Timing and Duration

	First	Average	Share
Intercept	0.4677***	0.7814***	0.2165***
	(0.0228)	(0.0116)	(0.0083)
Macro	-0.0225	-0.0375**	-0.0050
	(0.0318)	(0.0161)	(0.0116)
Female	0.0183	0.0154	0.0154*
	(0.0236)	(0.0120)	(0.0086)
Regular	-0.0384	-0.0876***	-0.0818***
	(0.0333)	(0.0169)	(0.0122)
Macro:Regular	-0.2448***	-0.0923***	0.0097
	(0.0462)	(0.0234)	(0.0169)
R^2	0.21	0.31	0.15
Adj. R^2	0.20	0.30	0.14
N	479	479	479

Note: First and *Average* refer to the first position and the average position of the interruptions. The position is calculated as the starting position of an utterance, in seconds, divided by the total duration of the presentation. *Share* refers to the total duration of audience utterances divided by the total duration of presenter and audience utterances. Gender is imputed from audio recordings. Standard errors are in parentheses.

Table 12: Total Interruptions

	Discussant		Regular
	Pre-Discussant	Total	Total
Intercept	1.7149***	10.8101***	11.5945***
	(0.4724)	(0.5358)	(1.1587)
Female	-0.2809	-0.0090	1.7218
	(0.6584)	(0.7468)	(1.3685)
Macro	2.2232***	0.8347	7.7591***
	(0.6326)	(0.7175)	(1.3715)
R^2	0.05	0.01	0.14
Adj. R^2	0.04	-0.00	0.13
N	252	252	227

Note: Pre-Discussant refers to the time in a presentation with a discussant but before the discussant presents, while "Total" refers to the entire paper presentation time including question and answer sections at the end. Gender is imputed from audio recordings.

Table 13: Speaker Probability of Sounding Happy at Time t Given Past Speaker Tones for Regular Format Talks

	Presenter^i|Audience ^{j}	Audience^i|Presenter^j
Intercept	0.179***	0.208***
	(0.025)	(0.025)
Macro	-0.003	-0.007
	(0.010)	(0.010)
Age^i	0.016*	0.007
	(0.009)	(0.005)
Female^i	0.216***	0.138***
	(0.020)	(0.017)
Age ^{j}	-0.005	0.019**
	(0.005)	(0.009)
Female ^{j}	-0.082***	-0.118***
	(0.017)	(0.021)
Happy$_{t-1}^j$	0.446***	0.389***
	(0.027)	(0.027)
Female^i × Female ^{j}	-0.069***	-0.024
	(0.022)	(0.023)
Female Share	-0.061**	0.005
	(0.029)	(0.029)
Other Lagged tones	Yes	Yes
R^2	0.23	0.18
N	2887	2955

Note: This table records how speaker happy tone depends on seminar characteristics and previous speaker characteristics for regular seminars. The column name "$X^i|Y^j$" corresponds to the regression estimating Equation 4 with speakers i and j. Tone for speaker i is measured from the first split of speaker i's utterance, and tone for speaker j is measured from the last five seconds of speaker j's utterance. The gender and age variables are imputed using our algorithms. * denotes $p < 0.05$, ** denotes $p < 0.01$, and *** denotes $p < 0.001$. The standard errors are OLS standard errors.

Table 14: Speaker Probability of Sounding Happy Given Past Speaker Tones for Discussant Format Talks

	Presenter^i|Audience ^{j}	Audience^i|Presenter^j
Intercept	0.194***	0.124***
	(0.024)	(0.027)
Macro	-0.015**	-0.007
	(0.007)	(0.008)
Age^i	0.017***	0.006
	(0.006)	(0.004)
Female^i	0.178***	0.117***
	(0.015)	(0.013)
Age ^{j}	-0.007**	0.010
	(0.004)	(0.007)
Female ^{j}	-0.061***	-0.135***
	(0.013)	(0.016)
Happy$^j_{t-1}$	0.438***	0.518***
	(0.026)	(0.029)
Female^i × Female ^{j}	-0.049***	-0.004
	(0.016)	(0.017)
Female Share	-0.059***	0.038*
	(0.021)	(0.022)
Other Lagged tones	Yes	Yes
R^2	0.2626	0.2514
Adj. R^2	0.2597	0.2486
N	2887	2955

Note: This table records how speaker happy tone depends on seminar characteristics and previous speaker characteristics for discussion seminars. The column name "$X^i|Y^j$" corresponds to the regression estimating Equation 4 with speakers i and j. Tone for speaker i is measured from the first split of speaker i's utterance, and tone for speaker j is measured from the last five seconds of speaker j's utterance. The gender and age variables are imputed using our algorithms. * denotes $p < 0.05$, ** denotes $p < 0.01$, and *** denotes $p < 0.001$. The standard errors are OLS standard errors.

Table 15: Count of Extreme Happy Tone Shifts from Presenter Characteristics

	No. most positive	No. most negative
Intercept	-0.0330	0.1267
	(0.1538)	(0.1521)
Female	0.6100**	0.3153
	(0.2405)	(0.2380)
Age	-0.0357	0.1282
	(0.1247)	(0.1234)
Macro	0.5278***	0.6748***
	(0.1462)	(0.1447)
Female Share	0.4257	0.4882
	(0.3694)	(0.3655)
R^2	0.19	0.18
Adj. R^2	0.17	0.17
N	227	227

Note: This table relates how many times a presenter experiences large swings in their happy tone to their traits and session traits for regular seminars only. The change in their happy tone is measured by subtracting the happy tone at the last minute of a presenter's first utterance by their happy tone at the first minute of their second utterance. The gender and age variables are imputed using our algorithms. * denotes $p < 0.05$, ** denotes $p < 0.01$, and *** denotes $p < 0.001$. The standard errors are OLS standard errors.

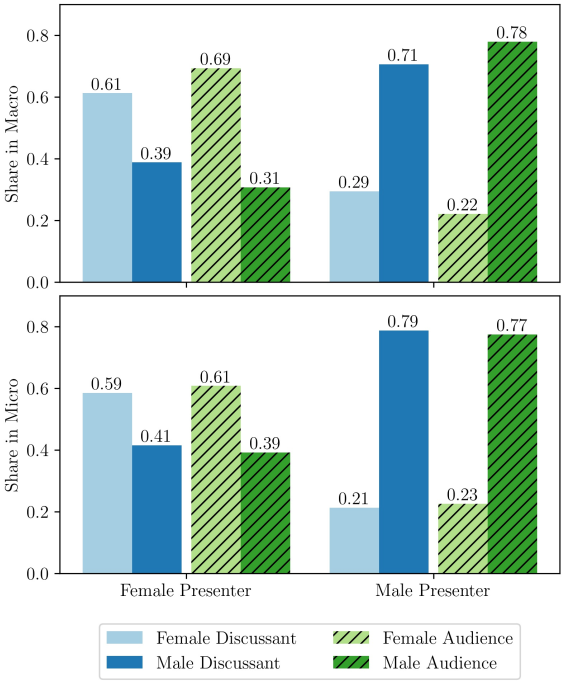
Table 16: Share of Interruptions with Extreme Happy Tone Shifts

	Share Extreme **Positive** Shifts	Share Extreme **Negative** Shifts
Intercept	0.008	0.008
	(0.008)	(0.009)
Female	0.047***	0.031**
	(0.013)	(0.014)
Age	0.000	0.004
	(0.007)	(0.007)
Macro	0.013*	0.029***
	(0.008)	(0.009)
Female Share	0.006	0.020
	(0.020)	(0.022)
R^2	0.18	0.15
Adj. R^2	0.16	0.14
N	227	227

Note: This table relates the share of interruptions a presenter experiences that dramatically shifts their probability of sounding happy. The change in their happy tone is measured by subtracting the happy tone at the last minute of a presenter's first utterance by their happy tone at the first minute of their second utterance. The gender and age variables are imputed using our algorithms. * denotes $p < 0.05$, ** denotes $p < 0.01$, and *** denotes $p < 0.001$. The standard errors are OLS standard errors.

B Figures

Figure 1: Discussant and Audience Gender Share



Note: Gender is imputed from audio data and our algorithm.

Figure 2: Confusion matrix for female audio clips, on model trained on male-only data

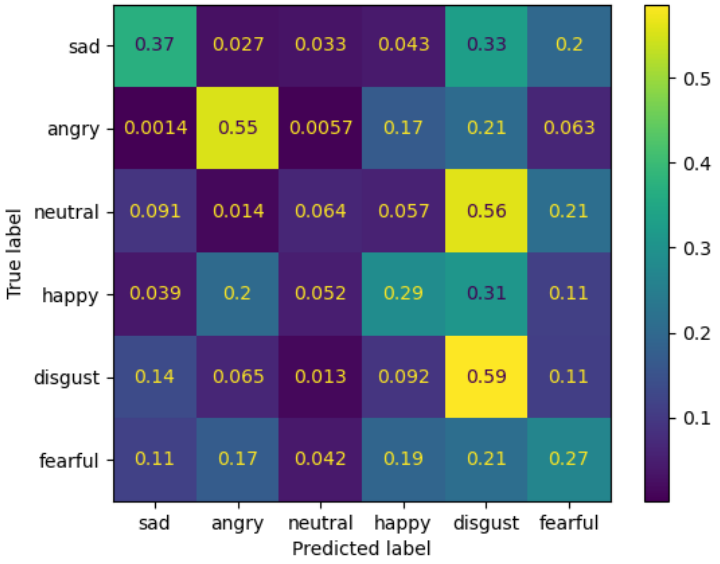
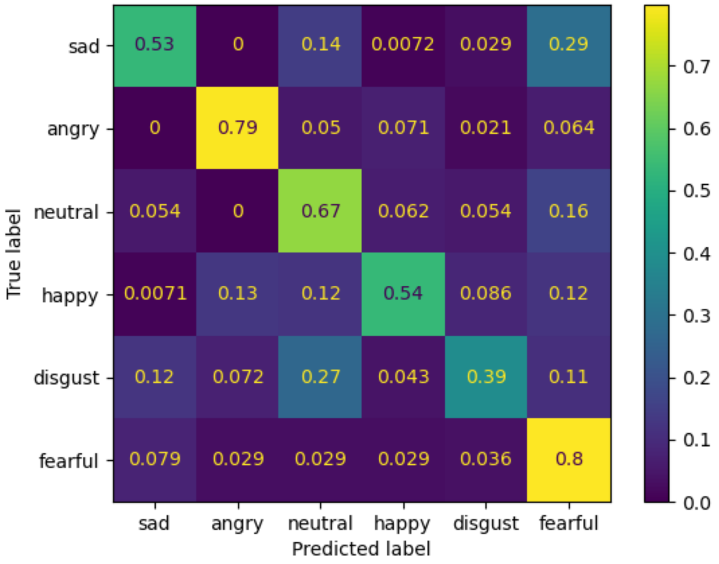


Figure 3: Validation confusion matrix for model trained on female-only data



Note: each row represents the distribution, in fractions, of predictions by true labels.