

PERSISTENCE OF VOICE PITCH BIAS AGAINST POLICY DIFFERENCES *

Asli Ceren Cinar[†] Özgür Kıbrıs[‡]

September 18, 2023

Abstract

Voters evaluate politicians not just on their policy positions and partisanship, but also on how they look and sound. But to what extent might these latter characteristics be traded for a more preferable policy stance? In this paper, we use an online experiment to study the relative effect on voter behavior of candidate voice pitch and policy differences between candidates. We first demonstrate a strong *voice pitch bias*: in an election between candidates who are identical in every aspect but voice pitch, voters are significantly more likely to choose the one with the lower voice pitch. Voice pitch bias is higher in elections between men than women candidates. We then introduce a novel phenomenon. *Persistence of voice pitch bias* is the amount of policy difference needed to compensate for voice pitch bias. While persistence is also gender-dependent, the effect is now reversed: voice pitch bias is more persistent in elections between women than men candidates. Finally, as a possible mechanism we show that voters perceive candidates with a lower voice pitch as more competent and trustworthy. Our findings are robust against voters' gender and other socio-economic characteristics.

Keywords: voice pitch bias, persistence, voting, perception, elections, online experiment

*We thank Fikret Adaman, Nejat Anbarcı, Abdurrahman Aydemir, Rachel Bernhard, Florian Foos, Sara Hobolt, Özge Kemahloğlu, Casey Klofstad, Korhan Kocak, Katharina Lawall, Mert Moral, and Emre Selçuk, as well as EPOP 2021, MPSA 2022, EPSA 2022, Oxford RAI Political Behaviour Workshop and LSE PSPE participants for comments and suggestions. The paper also immensely benefited from the inputs of the editor and two careful reviewers.

[†]Department of Government, London School of Economics; e-mail: a.cinar@lse.ac.uk

[‡]Faculty of Arts and Social Sciences, Sabancı University; e-mail: ozgur.kibris@sabanciuniv.edu.

1 Introduction

A growing body of literature has reliably established that political candidates benefit from having a lower voice pitch (as summarized in the next section). They are perceived to have higher leadership abilities. Hence, keeping everything else the same, a lower voice pitch brings in more votes. This *voice pitch bias* might be the reason why Margaret Thatcher, one of the most important political figures of the 20th century, worked with a voice coach early in her political career to lower her voice pitch.¹ This begs the question: what alternative strategies can a politician employ to overcome voice-pitch bias, and how effective are they?

A candidate’s policy position is arguably the most important aspect according to which she is evaluated by voters. Hence, by proposing a more desirable policy than her opponent, a candidate can be expected to offset the vote disadvantage due to voice-pitch bias. But in relation to one another, how important are voice-pitch and policy position? And how does the trade-off between the two depend on the candidate’s gender, the socioeconomic characteristics of the voter, and the policy dimensions on which the candidates contend? The objective of our study is to answer these questions using an experimental methodology. To the best of our knowledge, ours is the first study to approach these questions.

To analyze these issues, we propose a new measure that we call *persistence of voice pitch bias*: the amount of policy difference that makes a voter indifferent between two candidates who have a unit difference in voice pitch but are identical otherwise. More formally, persistence of voice pitch bias is the marginal rate of substitution of voice pitch for policy in the voter’s preferences.² Measuring its magnitude in an experimental setting, we find a *significant level of persistence in voters’ revealed preferences*. Furthermore, we find that *the voice pitch bias voters display is, on average, five times more persistent when evaluating women as opposed to men candidates*. Hence, a man candidate disadvantaged due to voice pitch bias needs a much smaller policy adjustment than a woman candidate in a similar situation. Our experiment also corroborates the earlier empirical and experimental literature discussed in the next section in that the *average voter exhibits a significant voice pitch bias*. Furthermore, we find that the *average voter displays a higher voice pitch bias when evaluating a*

¹Quoting Thatcher’s (1995, pg. 267) book, *The Path to Power*, “There was also the matter of my voice. ... an obvious problem for most women. ... my voice was naturally high-pitched, which can easily become grating. I had been told about this in earlier years and had deliberately tried to lower its tone. ... Gordon found me an expert who knew that the first thing to do was to get your breathing right, and then to speak not from the back of the throat but from the front of the mouth.”

²A voter whose utility depends on the policy choice (p) and voice pitch (v) of a candidate will have a utility function of $U(p, v)$ and the relative importance of the voice pitch will be captured in the marginal rate of substitution between the two: $\frac{\partial U / \partial v}{\partial U / \partial p} = \frac{\Delta p}{\Delta v}$. For a unit difference in the voice pitch, $\Delta v = 1$, persistence becomes *the amount of policy difference between the two candidates needed to offset the voter’s voice pitch bias*.

*man candidate than a woman candidate.*³

Our study has important implications for gender and politics. The female voice pitch typically ranges between 165 Hz to 255 Hz while the male voice has typically a much lower pitch, with a range from 85 Hz to 180 Hz. As a result, signals of social dominance and trust perceived from the voice pitch leverages men over women. Even when the two candidates are of the same gender, the effect of voice pitch on perception is prominent, for example favoring women candidates with a lower voice pitch. Elections between men candidates suffer from a similar gender bias, and as discussed above, sometimes to a higher extent. Furthermore, we theorize that policy declarations by men candidates are taken much more seriously by voters, and that, a combination of these two biases results in our two main findings, namely that (i) voters exhibit a stronger voice pitch bias when evaluating men candidates, and yet (ii) voters respond more strongly to policy differences between men candidates, making voice pitch bias more persistent for women candidates.

We also analyze how participant gender affects voting behavior. We start with the case where there are no policy differences between candidates. For men participants, we find that a higher percentage vote for the lower-pitch candidate when choosing between men than women candidates. This is not the case for women participants. Hence, we conclude that it is men participants that drive the overall finding that voice pitch bias is higher for men candidates than women candidates. On the other hand, both men and women participants strongly respond to a small policy difference between men candidates, but not between women candidates. Additionally, we analyze the implications of a variety of pre-treatment covariates⁴ and demonstrate that our findings are robust against their inclusion in the analysis.

As two potential mechanisms for voice pitch bias and its persistence, we also measure how a candidate’s voice pitch affects the participants’ perception of her/his competence and trustworthiness. Competence is closely related to social dominance and the literature discussed in the next section shows that a candidate with lower voice pitch is perceived as more competent. On the other hand, previous findings on trustworthiness, another important characteristic for a candidate, are mixed. We contribute to this discussion by showing that

³In an election between a lower-pitch (LP) and a higher-pitch (HP) candidate who propose identical policies, a voter’s *voice pitch bias* can be measured via the likelihood that they will vote for the LP candidate. For an unbiased voter, this probability would be 50%, since the only difference between the two candidates – the voice pitch – would be irrelevant for them. If, however, the voter is more likely to vote for the LP candidate than can be attributed to chance, the difference in probability is a measure of the voter’s voice pitch bias.

⁴We consider age, household income, whether the participant voted in the last election, left-right ideology, trust towards others, level of satisfaction from existing education and health policies, importance of government spending for education and health, views on whether men or women in elected office are better at handling issues with regarding to education or health, as well as participants’ completion times.

a lower voice pitch creates on our participants a perception of both higher competence and higher trustworthiness. Furthermore, the effect is stronger for voters evaluating men candidates. In Section 4 we discuss how these findings might provide an underlying mechanism for voice pitch bias.

We focus on two policy dimensions, namely, per capita *public spending on health* and *public spending on education*, varied in a between-subjects design. Focusing on per capita public spending allows policy differences between candidates to be measured in monetary terms. We choose health and education since both are major valence issues, i.e., issues where there is a broad consensus that an increase in the current level of public spending is desirable in Turkey, where we run our experiment, and both have been salient policy dimensions around the world, especially after the onset of Covid-19. Additionally, the Turkish public is predominantly satisfied (dissatisfied) from government policies on health (education), hence allowing us to control for attitudes towards government.

In our experiment, participants listen to two voice recordings (*i.e.*, candidates). On a given policy dimension, each recording declares a policy stance (*e.g.*, I will annually allocate X TL⁵ per person for public health expenditures). To control for unobservable individual differences between candidates, the two voice recordings are obtained from the same person and digitally manipulated to either a higher or a lower pitch to create a difference of 1 Equivalent Rectangular Bandwidth (ERB) (roughly 40 Hz) between them. We refer to these two recordings as the *lower-pitch* (LP) and the *higher-pitch* (HP) candidates.⁶ In our design, a participant is exposed to candidates of only one gender. Our experiments were carried out online. Online voice pitch experiments have been shown to produce results that are comparable to laboratory experiments [Feinberg et al., 2008].

2 Background and Hypotheses

Voice pitch is a prominent vocal feature that has significant effect on humans’ perception of others. It can be defined as the number of vibrations per second made by the vocal folds to produce a vocalization [Tusing and Dillard, 2000]. Larger vocal folds generate lower frequencies due to slower vibrations, and hence produce lower sounding voices. Additionally, the human voice pitch is sexually dimorphic [Puts et al., 2007]. The voice pitch of an average male is almost half of that of an average female [Titze, 1994, Feinberg et al., 2005a, Vieira et al., 2015]. The discernible dimorphism can be attributed to factors beyond mere

⁵We use Turkish Liras (TL) as our experimental currency. At the time of the experiment 6.9 TL was equal to 1 USD.

⁶In line with the literature (*e.g.* Klofstad et al. [2012] and Tigue et al. [2012]) and as confirmed with our validation tests, the voice pitch difference between the two recordings is sufficient to create on the listener the impression that the two voices do not belong to the same person.

discrepancies in physical dimensions between genders. It has been observed that men tend to have a lower vocal pitch in comparison to women and prepubescent children of both sexes when taking into account their respective height and body volume [Titze, 2000]. According to the literature, the emergence of sexual dimorphism in the human voice can be attributed to sexual selection, specifically through the process of female mate choice [Darwin, 1871, Collins, 2000]. The literature shows that the voice pitch is perceived to signal information about physical and psychological traits such as attractiveness [Collins, 2000, Collins and Missing, 2003, Feinberg et al., 2005a,b, Jones et al., 2008], social and physical dominance [Tigue et al., 2012, Puts et al., 2007, Rezlescu et al., 2015, Schild et al., 2022], and reproductive capabilities [Feinberg et al., 2005b]. Lower voices are perceived to signal masculinity, trustworthiness, competence, and strength [Puts et al., 2007, O’Connor and Barclay, 2017, Klofstad, 2016, 2017, Banai et al., 2017, Feinberg et al., 2008, 2005b, Jones et al., 2010]. Hence, in environments where these traits are desirable listeners are expected to exhibit a bias against higher-pitched speakers.

In the context of politics, a number of experiments demonstrate that candidates with lower voices receive higher votes and have a higher probability of winning elections [Anderson and Klofstad, 2012, Tigue et al., 2012, Klofstad et al., 2012, 2015, Klofstad, 2016].⁷ Banai et al. [2017] supports these findings in an empirical study of 51 presidential elections around the world. Klofstad [2016] analyzes the 2012 US House Elections and provides an overall support with an exception that will be discussed below. Some studies also analyze how voice pitch bias interacts with other variables. Laustsen et al. [2015] finds in survey experiments that voters with a more conservative stance display a higher voice pitch bias than more liberal voters. Using empirical data as well as survey experiments, Klofstad [2016] shows that older, well-educated, and politically engaged voters are the most biased in favor of candidates with lower voices. In an experimental study, Klofstad et al. [2015] establishes candidate age as an important determinant of voter choice and shows voice pitch to have an effect on perception of candidate age. In an empirical study, Klofstad and Anderson [2018] finds no correlation between a politician’s voice pitch and leadership ability.

The first main contribution of our paper is to this literature. First, we replicate the aforementioned finding that between two candidates who are identical in every aspect but their voice pitch, a voter is more likely to vote for the one with the lower voice.⁸

More importantly, in a novel experimental design that has not been considered before, we differentiate the two candidates on the policy space and study how voice pitch bias interacts with such differences. Specifically, we measure how much of a policy difference between the

⁷Similar results prevail outside the context of politics, for example in case of job applications [Ko et al., 2009].

⁸To the best of our knowledge, our study is the first to replicate this important finding in a non-English speaking context.

LP and HP candidates is sufficient to offset voice pitch bias. As discussed in Section 1, this amount gives us an estimate of the voter’s marginal rate of substitution between voice pitch and policy. It is what we refer to as *persistence of voice pitch bias*.

Hypothesis 1. *By proposing a more desirable policy than her (his) opponent, a candidate can offset the vote disadvantage due to voice pitch bias.*

Gender is one of the most prominent traits voters perceive at first sight. Hence, its effects on politics have been the subject of extensive research. The literature indicates that voters believe women politicians to be warmer, more compassionate, better able to handle education, family, and women’s issues, more liberal, and feminist than men, whereas men politicians are seen as strong, intelligent, better suited to handle crime, defense, and foreign policy issues, and more conservative (see Johns and Shephard [2007], Dolan [2010] and the literature cited therein). These gender stereotypes affect voting behavior significantly. Koch [2000] analyzes data from the 1988–1992 Pooled Senate Election Study to show that even after candidates’ individuating ideological orientations are taken into account, candidate gender still exerts substantial effect on how voters perceive a candidate’s ideological orientation. Eagly et al. [2003] shows that voters are more likely to vote for candidates who endorses a position typically favored more by their own gender. Dolan [2010] examines how gender stereotypes shape voters’ support for women candidates in various electoral circumstances. Johns and Shephard [2007] finds that men voters are more inclined than women to see men candidates as stronger and to prioritise strength while voting. Bernhard [2023] finds that women candidates significantly benefit from being taller than their opponents while the benefit of height is not significant for men.

Mo [2015] analyzes how candidate quality and voter gender bias interact to determine candidate evaluation. Relatedly, Bauer [2020] finds that voters hold women politicians to higher qualifying requirements than men. These higher standards make it more difficult for women candidates to acquire electoral support. Hence, women candidates on average are more qualified than their men counterparts. Furthermore, Fox and Lawless [2004] uncover that, on average, women, even those with the highest levels of professional achievement, are less likely than men to consider running for political office.

In general, women candidates face gendered constraints when running for office, and are required to “double-bind” themselves by demonstrating the competence associated with masculinity and the tenderness associated with femininity [Carpinella and Bauer, 2021, Bauer and Santia, 2021]. Relatedly, Schneider and Bos [2014] find that women politicians do not possess the traits attributed to women (*e.g.*, warm, empathetic), and they have no advantage in terms of female-stereotypical characteristics. Dietrich et al. [2019] quantify lawmakers’ emotional intensity by analysing minor voice pitch fluctuations. They find that women display greater emotional intensity when talking about women related issues than about other

topics, and compared to men colleagues. Using role congruity expectations as a framework, [Boussalis et al. \[2021\]](#) examines how candidate gender affects usage of facial, vocal, and textual communication in German federal election debates (2005–2017), as well as voters’ reaction to such communication. For example, they find that Angela Merkel expresses less anger than her male opponents, and that voters punish her for anger displays and reward her for happiness and general emotional displays. [Carpinella and Bauer \[2021\]](#) demonstrate that women candidates tend to blend male verbal assertions with feminine images such as presence of family, schools and hospitals.

Regarding the effect of voice pitch, [Searles et al. \[2017\]](#) analyzes relative effectiveness of man and woman voices in political advertising, in relation to the considered issues being masculine or feminine. Candidate and voter gender turns out to be important for voice pitch bias as well. [Klofstad \[2016\]](#) analyzes 2012 data on the US House Elections and finds that when facing a woman opponent, a higher voice pitch increases votes. [Anderson and Klofstad \[2012\]](#) finds that when considering men candidates for feminine leadership roles, women voters do not respond to voice pitch (while man voters do).

The second main contribution of our paper is to the literature on the effect of gender on voter preferences, summarized above. We first test whether the amount of voice pitch bias depends on the candidate gender. More importantly, we also test whether candidate gender affects the marginal rate of substitution between voice pitch and policy. We hypothesize that the effect of voice pitch is more persistent in case of women candidates and hence, to offset voice pitch bias an HP woman needs to offer a much more desirable policy than an HP man.

Hypothesis 2. *Voters exhibit a more persistent voice pitch bias when voting between women candidates than voting between men candidates.*

The literature shows voice pitch to have a strong influence on the perception of characteristics related to social power, such as competence or social dominance (*e.g.*, see [Aung and Puts \[2020\]](#)). Individuals with lower voices are perceived to be more competent [[Klofstad et al., 2015](#)], more socially dominant [[Tigue et al., 2012](#), [Gregory, 1994](#), [Wolff and Puts, 2010](#), [Borkowska and Pawlowski, 2011](#), [Jones et al., 2010](#), [Klofstad et al., 2015](#), [Ko et al., 2009](#), [Puts et al., 2007](#), [Laustsen et al., 2015](#)], and have better leadership abilities [[Nagel et al., 2012](#), [Klofstad et al., 2015](#)]. The effect of voice pitch on perceptions of trustworthiness is, on the other hand, gender-dependent. [O’Connor and Barclay \[2017\]](#) and [Klofstad et al. \[2012\]](#) find in two alternative contexts that lower pitch women voices are perceived to be more trustworthy. For man voices on the other hand, [O’Connor and Barclay \[2017\]](#) find that a higher pitch induces more trust, while [Tigue et al. \[2012\]](#) obtain an opposite finding. These contrasting findings present an interesting puzzle for us to focus on.

Our paper also contributes to this discussion by measuring voters’ perceptions of competence and trustworthiness of both men and women candidates when they are recorded

making a policy-neutral statement “Vote for me”. In case of trustworthiness, our paper also serves to bring further evidence to the contrasting findings in the earlier literature.

3 Materials and Methods

3.1 Experimental Stimuli

We recorded six native Turkish speakers - three females with an average age of 38 and three males with an average age of 40 - making the following policy statements in Turkish: ‘Please vote for me’ and ‘I will annually allocate X TL per person’. [Pisanski et al. \[2021\]](#) shows that studies on voice pitch obtain comparable results over different types of recordings, such as a series of vowels, a single word, or a sentence as in our case. We recorded multiple speakers to reduce any individual-level effect of other vocal characteristics such as tone of the voice, rhythm, or tempo. The monetary amount X took six values, starting from 10,000 Turkish Liras (TL) and decreasing by 200 TL at each step down to 9,000 TL. Overall, we obtained 7 recordings from each speaker.

We used monetary differences in per capita public spending to measure differences in policy. This is because monetary differences in per capita spending are easy to understand and their perception is uniform among participants (as opposed to for example differences in an abstract policy space). Also, they allow us to measure in objective units the persistence of voice pitch bias (the trade-off between voice pitch bias and policy differences).

As discussed in the previous section, we expect to corroborate the previous literature by finding that the LP candidate will receive significantly higher votes than the HP candidate. We then hypothesize that policy differences between candidates can first mitigate and then neutralize this voice pitch bias, at which point the percentage of participants voting for the LP candidate will not be significantly different than 50% (see Hypothesis 1). To measure how much of a policy difference is needed to neutralize the voice pitch bias, we gradually made the policy declaration of the low-pitched candidate less desirable. After analyzing pre-test results, we determined these policy increments to be 200 TL.

To reduce any policy-specific effect on our findings, we chose two alternative policy dimensions. Our main objective in choosing education and health was to find valence issues where a higher public spending would be almost unanimously considered to be desirable. [Figure A.2](#) in the Appendix shows that this is indeed the case for our participants.⁹ Another reason for our choice of education and health was that the Turkish public is predominantly satisfied (dissatisfied) from government policies on health (education), hence allowing us to

⁹All tables and figures on our descriptive statistics are in the Appendix.

control for attitudes towards government.¹⁰ This is indeed corroborated by our data, as displayed in Figure A.3 in the Appendix. Hence, any finding that holds for both dimensions can not be attributed to the participants’ attitude towards government. Finally, in a post-experiment survey, we asked the participants whether they think it is men or women in elected office that are better at handling each issue. As can be seen in Table 1 in the Appendix, more than 75% of our participants express no preference and the remaining group is equally divided between preference for men and women. Hence, we believe it is appropriate for our study to consider education and health as gender-neutral issues.

To find a common monetary unit for both education and health policy statements, we consulted the OECD education and health reports for Turkey where the yearly per capita spending is stated to be approximately 2,400 TL for both dimensions. We then chose our maximal amount (10,000 TL) to be significantly higher.

Voices were recorded as .mp4 files.¹¹ We inspected each audio file aurally and visually in Audacity (v.2.3.3).¹² Before converting the audio files into .wav format, we ensured that the recordings were without speech errors and background noise. We used the *Get Pitch* command in the Praat phonetic analysis program [Boersma and Weenink, 2020, v.6.1.15] to determine the mean pitch of each recording. For unaltered woman voices, the mean pitch is 239 Hz and the standard deviation is 14 Hz. The mean pitch for unaltered man voices is 134 Hz and the standard deviation is 12 Hz. To create a lower-pitched and higher-pitched version of each recording, we used the Pitch-Synchronous Overlap Add (PSOLA) method in Praat.¹³ Following the literature (e.g. see Jones et al. [2010], Klofstad and Anderson [2018], Tigue et al. [2012]), we altered each recording by ± 0.5 Equivalent Rectangular Bandwidth (ERB).¹⁴ Hence, each recording was converted into a pair of recordings, one with a higher-

¹⁰For example, see the 2020 Life Satisfaction Survey by the Turkish Statistical Institute.

¹¹The audio recordings were captured using a smartphone microphone, specifically an iPhone 11, in a quiet office environment. This was necessitated by the constraints imposed on data collection during the COVID-19 lockdown period. Existing literature suggests that while an environment with sound treatment is optimal for collecting voice samples, a microphone on a smartphone can still obtain recordings that are suitable for the analysis of acoustic signals, especially fundamental frequency (f_o) [Van der Woerd et al., 2020, Uloza et al., 2021, Fahed et al., 2022]. More specifically, Van der Woerd et al. [2020] results indicate that there are no significant main effects or interactions observed for the average f_o across four recording conditions: (i) recording with an iPhone in an office setting, (ii) recording with a BlueYeti USB microphone in an office setting, (iii) recording with an iPhone in a sound-proof chamber setting, and (iv) recording with a BlueYeti USB microphone in a sound-proof chamber setting.

¹²Audacity® software is copyright © 1999-2020 Audacity Team. The name Audacity® is a registered trademark of Dominic Mazzoni.

¹³This method allows us to manipulate fundamental frequency and harmonics whilst controlling the other spectrotemporal aspects of the acoustic signal [Feinberg et al., 2005a].

¹⁴In determining this magnitude, there is a trade-off between obtaining natural sounding voices – which favors smaller alterations – and creating the perception that the two voices are different – which favors bigger

pitch and one with a lower-pitch with 1 ERB difference between the two. The ± 0.5 ERB manipulation creates natural sounding voices and accounts for a perceivable shift of roughly ± 20 Hz. Manipulating the recordings by ERB corrects for the logarithmic difference between actual fundamental frequency and perceived fundamental frequency. Therefore, it produces a constant perceivable gap between the raised and lowered versions of a recording, regardless of its initial fundamental frequency. As explained in the Appendix, this is confirmed for our experiment as well.

3.2 Procedure

The experiment was carried out online, using the software Qualtrics. Results obtained from online voice pitch experiments have been shown to be comparable to those of laboratory experiments [Feinberg et al., 2008]. As summarized in Part (a) of Figure 1, the experimental conditions were assigned following a 2×2 factorial design (with equal probability): the participants were randomly assigned to listen to policies about either only education or only health and the candidates they were evaluating in-between were either always both men or women. Participants compared recordings of multiple speakers in order to minimize pseudoreplication bias, whereby the idiosyncratic characteristics of any one speaker might influence the results of the experiment [Kroodsma, 1990, Machlis et al., 1985]. The experiment was not pre-registered.

In the first part of the experiment (Figure 1, first step of Part (a)), a participant listened to 6 individual recordings (obtained from 3 speakers) speaking the sentence ‘Please vote for me’. After listening to each recording, the participant rated the candidate in terms of trustworthiness and competence.¹⁵

In the second part of the experiment (Figure 1, second step of Part (b)), a participant listened to 18 pairs of recordings obtained from 3 different speakers. After listening to each pair, the participant was asked to vote for one of the candidates (Figure A.1 in the Appendix presents a screenshot of a typical choice task). In the health policy treatment, each recording declared ‘I will annually allocate X TL per person for public health expenditures’. In each recording-pair, both recordings were obtained from the same speaker but one was higher-pitched (HP candidate) and the other one, lower-pitched (LP candidate). While the HP recording always stated $X = 10,000$ TL, the LP recording declared an X in between 9,000 TL and 10,000 TL with increments of 200 TL. The order of recording pairs, as well as, the order of recordings in each pair, were determined randomly to eliminate order effects.

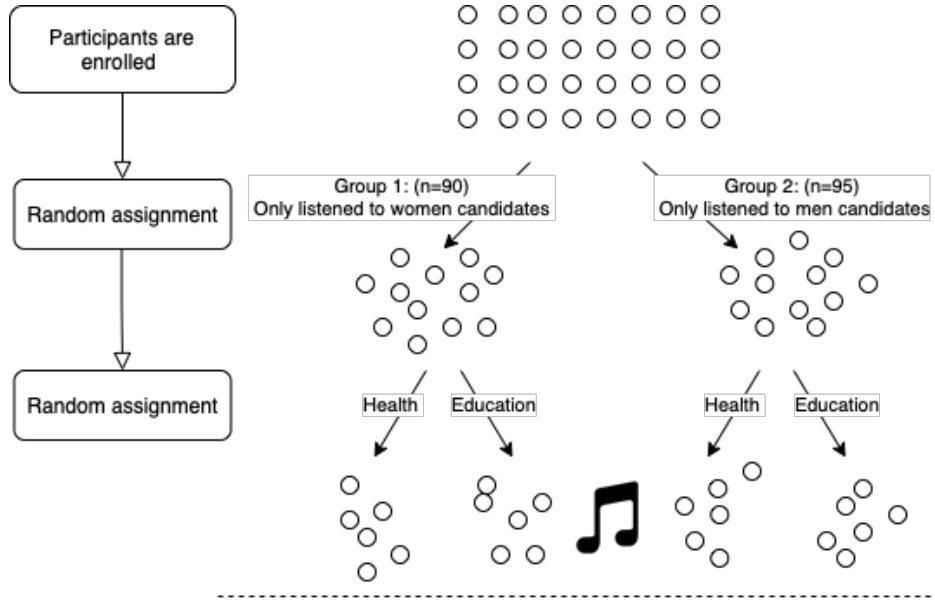
We did not inform the participants about the gender of the candidates they are listening to. Hence, any differentiation between two participants evaluating candidates of different

alterations. The previous literature suggests a ± 0.5 ERB alteration to provide a good trade-off.

¹⁵Participants chose whether to use computer speakers ($n = 142$) or headphones ($n = 45$).

Figure 1: Experimental Conditions

(a) Random assignment of experimental conditions (between-subject)



(b) Decision tasks for each participant (within-subject, random order)

First step:

HP: "Please vote for me" vs LP: "Please vote for me"

Second step:

HP: 10,000 TL	vs	LP: 10,000 TL
HP: 10,000 TL	vs	LP: 9,800 TL
HP: 10,000 TL	vs	LP: 9,600 TL
HP: 10,000 TL	vs	LP: 9,400 TL
HP: 10,000 TL	vs	LP: 9,200 TL
HP: 10,000 TL	vs	LP: 9,000 TL

gender is solely based on these participants’ perceived gender norms regarding women and men, particularly about their voice pitch.

In the third and final part of the experiment, participants filled out a survey including questions on their birth year, sex, gender, education, family income, whether they voted in the last elections, political preferences, trust towards others, opinion on whether men or women politicians are better suited for public health/education issues, satisfaction level from public health/education services, and importance of government providing public health/education services. Participants were also asked if they had trouble listening to the recordings (Table 4 in Appendix) and the medium they used for listening (speakers versus earphones). Covariates on gender, age, ideology, income, turnout in the last general election, survey completion time, listening device (mobile phone versus computer), medium of listening (earphones versus speakers), and general trust show no significant differences across experimental groups (See Tables 3, 4 and 5 in the Appendix).

Participants

Participants ($N = 185$) predominantly declared their gender as either woman (68 participants) or man (113 participants).^{16,17} Participants ranged in age from 19 to 29 (mean age of the participants were 22 with a standard deviation of 0.12). Students who were enrolled in the introductory courses of economics received an email link that gave them access to the survey in the Qualtrics platform. Participants received course credit in exchange for their participation. They came from a diverse range of majors, belonging to the faculties of engineering, management, and social sciences. Anonymity of the participants was respected throughout the study, and their identities were kept confidential.¹⁸

Additional socioeconomic characteristics of our sample are as follows. Prior to our study, 93% of participants had voted in a real-life election. This is not significantly different than the 86% turnout rate in the 2018 Turkish elections, as published by the Turkish Statistical Institute. The monthly median family income in our sample is between 15,000 TL (2,170 USD) and 18,000 TL (2,600 USD), with a standard deviation of 10,800 TL (1,565 USD). The ideological distribution of our participants, another covariate that we are interested in, accumulates toward the mid-point, as seen in Table 1 (Appendix). Furthermore, as noted in Subsection 3.1, our participants consider a higher public spending on both health and education to be desirable (Figure A.2, Appendix). Additionally, they are predominantly

¹⁶3 participants preferred not to answer and 1 participant declared themselves queer.

¹⁷Protocols for this study were approved by Sabancı University Ethics Committee with protocol number FASS-2020-08.

¹⁸To be able to give bonus credits, course instructors were given the identification numbers of the students who participated in the study. But they did not have access to any other data.

satisfied with public health services while the satisfaction numbers are significantly lower for education (Figure A.3, Appendix). Tables 1 and 2 in the Appendix respectively present descriptive statistics for each experimental group and results of balance tests.

4 Results

4.1 Voice Pitch Bias

We summarize the results of our experiment in Figures 2 and 3. For each level of policy difference (taking values from 0 to 1000 on the x -axis), we conducted separate linear regressions where the randomly assigned candidate gender is the independent variable. For each participant, the dependent variable is her average vote for the LP candidate, taken over the three choice tasks (recordings).

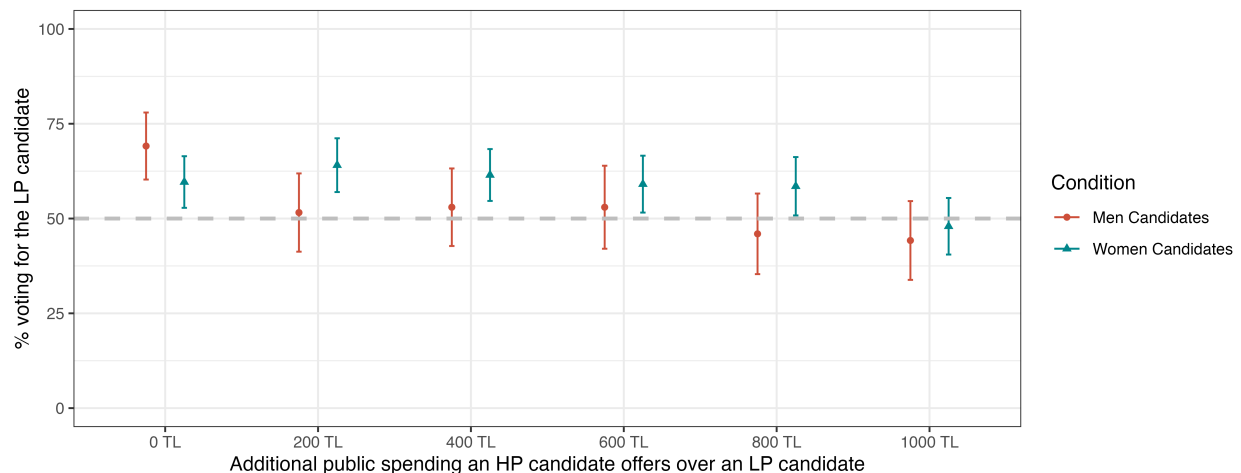


Figure 2: Percentage of participants who voted for the LP candidate by treatment condition, 95% CIs.

Figure 2 displays – for both men and women candidates – how the percentage of participants voting for an LP candidate changes in response to the policy difference between LP and HP candidates. For both LP men and LP women, a policy difference of no more than 1,000 TL is sufficient to take this percentage down to a neighborhood of 50%, providing support for Hypothesis 1. We hence conclude that by offering a more desirable policy than an LP opponent, an HP candidate can increase their vote shares and offset voice pitch bias. However, the amount of policy difference needed depends on the candidate gender, as discussed in the next paragraph.

Figure 2 shows that the percentage of participants voting for an LP man displays a sharp decrease to almost 50% as the policy difference between the candidates increases to 200 TL

($p > 0.1$).¹⁹ That is, by offering 200 TL more public spending than the LP man, the HP man is able to offset the vote disadvantages of voice pitch bias. This observation is significantly different from what we see in elections between women candidates. As can be seen in Figure 2, the percentage of participants voting for an LP woman remains significantly higher than 50% ($p < 0.05$) as the policy difference increases from 0 up to 800 TL. That is, even by offering an 800 TL more favorable policy than her opponent, an HP woman can not offset the detrimental effect of voice pitch bias on her votes. It is only when the policy difference reaches 1,000 TL that an LP and an HP woman receive more or less the same number of votes ($p > 0.1$). Overall, a comparison of men and women candidates shows us that voters exhibit a more persistent voice pitch bias when voting between women candidates than between men candidates. Hence, our findings support Hypothesis 2.

Figure 2 also provides support for the earlier literature by showing that *voters exhibit a significant voice pitch bias*: when there is no policy difference between the LP and HP candidates, the percentage of participants voting for the LP candidate is 64.50%, which is significantly higher than 50% ($p < 0.01$). We also analyze subsamples where all participants vote between either all men or all women candidates. We find the percentage of participants voting for an LP man to be 69.12% ($p < 0.01$), in comparison to 59.63% for an LP woman ($p < 0.01$). Hence, in both experimental groups the probabilities are significantly higher than 50%.

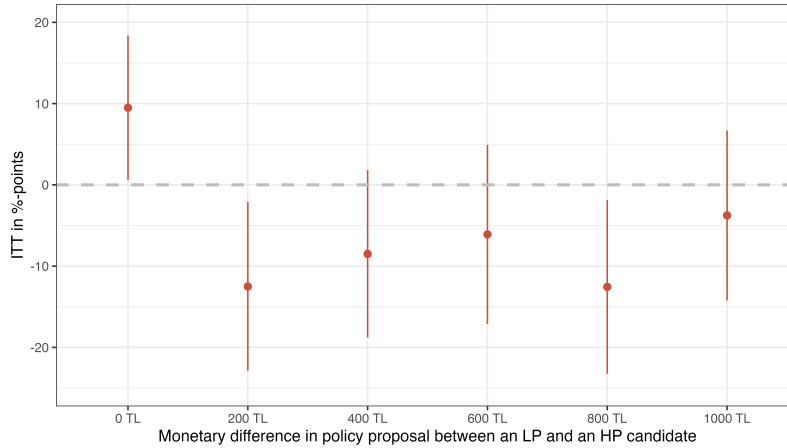


Figure 3: Effect of policy differences on the vote difference between an LP man and an LP woman.

Finally, Figures 2 and 3 together show that *the magnitude of voice pitch bias responds to candidate gender*. In Figure 3, when the horizontal axis is 0 TL (that is, when both

¹⁹This p-value belongs to the one-sided t-test for the significance of the difference between the percentage votes an LP man receives at 200 TL and 50%. The other p-values in this section have a similar interpretation.

candidates offer the same policy) we see a significant ($p < 0.05$) *Intention-To-Treat* (hereafter, *ITT*) effect of 9.49 percentage points. This means that the difference in votes received by an LP man and an LP woman is 9.49 percentage points. Hence, we conclude that voters exhibit a higher voice pitch bias when voting between men candidates than voting between women candidates. Figure 3 also shows that the effect is reversed in case of a 200 TL policy difference. That is, an LP man receives 12.49 percentage points less votes than an LP woman ($p < 0.05$). This is because an HP man offering 200 TL more than his LP opponent can overcome voice-pitch bias – decreasing his opponent’s votes down to approximately 50% – while an HP woman can not, hence demonstrating Hypothesis 2.

4.2 Perceptions of Trustworthiness and Competence

Figure 4 displays the effect of a switch in voice pitch from HP to LP on perceptions of competence and trustworthiness. For candidates of each gender, we conducted separate linear regressions with the candidate voice pitch as a binary independent variable and, for each participant, her average competence (respectively trustworthiness) evaluation as a dependent variable.

Over all experimental groups, the effect on a participant’s competence rating of a switch from an HP to an LP version of the same recording is 9.43 percentage points ($p < 0.01$). Hence, perceptions of competence provide a possible mechanism for the effect of voice pitch on voting behavior. In the two experimental groups – participants who only listened to women candidates versus participants who only listened to men candidates – the effect is 6.20 percentage points ($p < 0.05$) in case of men candidates and 12.84 percentage points ($p < 0.01$) in case of women candidates.

We next analyze another theoretical mediator, trustworthiness. Figure 4 shows that, over all experimental groups, the effect on a voter’s trustworthiness rating of a switch from an HP to an LP version of the same recording is 5.13 percentage points ($p < 0.05$). However, for reasons that will be discussed next, trustworthiness perception seems to be a less likely mechanism underlying voice pitch bias. Namely, when the two experimental groups – participants who only listened to women candidates versus participants who only listened to men candidates – are analyzed separately, the effect on trustworthiness ratings largely diminish in significance. In case of men candidates the effect is 4.34 percentage points and not significant ($p > 0.1$). In case of women candidates, the effect is slightly larger with 5.97 percentage points, and only significant with 90% confidence ($p < 0.1$).

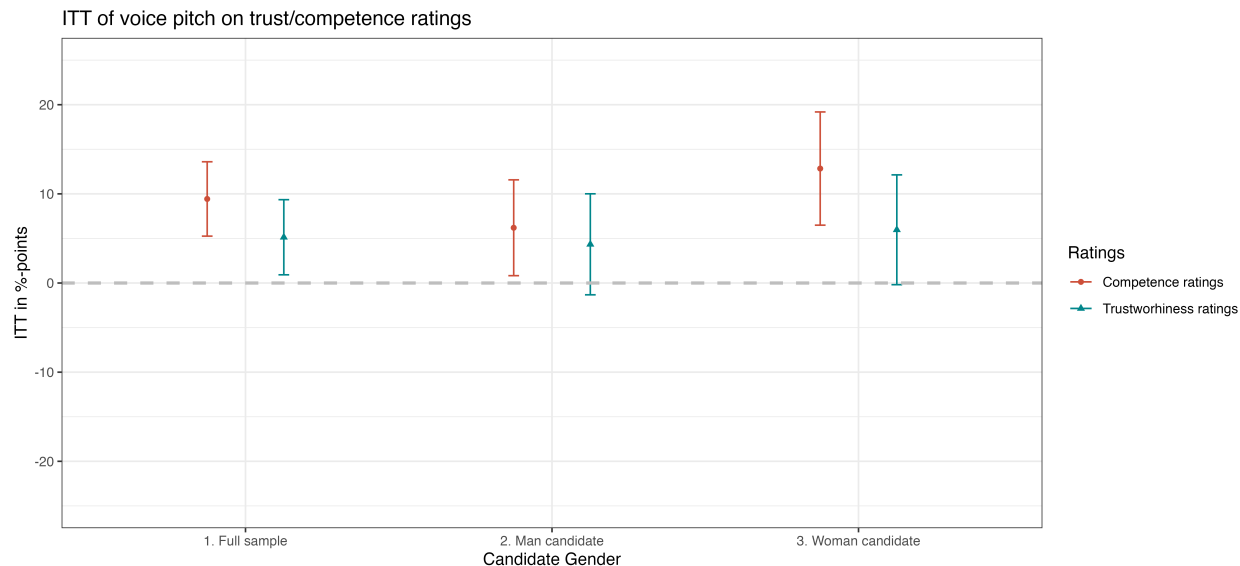


Figure 4: The effect of a switch in voice pitch from HP to LP on perceptions of competence and trustworthiness, 95% CIs.

4.3 Exploratory Analysis: Participant Gender

In this section, we demonstrate that participant gender is an important factor in our findings on candidate gender that (i) voters exhibit a higher voice pitch bias when evaluating men candidates but (ii) they exhibit a more persistent voice pitch bias when evaluating women candidates. Figure 5 displays the effect of participant gender on voting behavior. We first look at the case when the policy difference between candidates is 0 TL. We find that men participants vote 18.3 percentage points more for an LP candidate in elections between men than women ($p < 0.01$). In contrast, women participants vote 3.7 percentage points *less* for an LP candidate in elections between men than women, though the difference is not significant ($p > 0.1$). Hence, we conclude that it is men participants that drive the overall finding that voice pitch bias is higher for men candidates than for women candidates.

For comparison, in Figure 5 we also present the case when the policy difference between candidates is 200 TL.²⁰ Contrary to the previous case, we now find a similar treatment effect for both men and women participants. More specifically, women participants vote for an LP candidate 17.40 percentage points less in elections between men than women ($p < 0.05$). Similarly, men participants vote for an LP candidate 8.46 percentage points less in elections between men than women, though the difference is not significant ($p > 0.1$). Together with Figure 2, the above analysis suggests that for participants of both genders, a 200 TL difference between the candidates' policies results in a significant decrease in the probability

²⁰The other cases where the policy difference is higher look very similar to the case of 200 TL. Hence, for the sake of presentation we do not show those in Figure 5.

of voting for an LP man, providing a mechanism for Hypothesis 2.

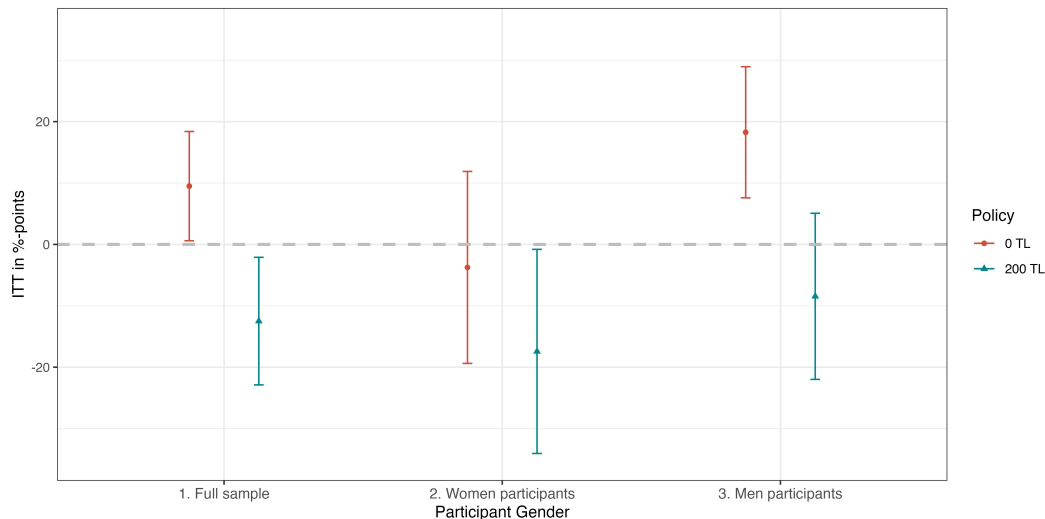


Figure 5: Effect of listening to men versus women candidates conditional on participant gender, 95% CIs.

5 Conclusion

By and large, our findings corroborate the stated hypotheses as well as the earlier literature. Our participants exhibit both voice pitch bias and persistence of voice pitch bias (Hypothesis 1). Furthermore, voice pitch bias is higher in elections between men candidates while it is more persistent in elections between women candidates (Hypothesis 2). We find that men participants are mainly responsible for gender dependence of voice pitch bias while both men and women participants drive gender dependence of the persistence of voice pitch bias. We also identify the effect of voice pitch on perceptions of competence and trustworthiness as an important mechanism for voice pitch bias.

Our results are robust to the inclusion of pre-treatment covariates, including participant gender, age, income level, turnout, ideology, general trust towards others, as well as survey completion time, which is a proxy for participant attention, listening device, and medium of listening. Additionally, our findings are consistent across the two policy dimensions – health and education – that we considered. Since the Turkish public is predominantly satisfied (dissatisfied) from government policies on health (education), we conclude that our findings are not driven by participants’ attitude towards government.

While we find that candidates with a lower voice pitch are perceived to be both more competent and trustworthy, our results also display an interesting pattern. As can be seen

in Figure 4, the effect of voice pitch on perception is higher for voters evaluating women candidates. This difference is particularly pronounced in competence ratings. Even though our participants overwhelmingly declare both education and healthcare to be gender-neutral issues, the earlier literature establishes them to be women-congruent [Dolan, 2014]. In relation to this literature, our study then shows that, on woman-congruent issues, even a vocal characteristic signaling masculinity – voice pitch – can have a more significant effect on perceptions regarding women than men.

References

- R. C. Anderson and C. A. Klostad. Preference for Leaders with Masculine Voices Holds in the Case of Feminine Leadership Roles. *PLoS ONE*, 7(12):e51216, Dec. 2012.
- T. Aung and D. Puts. Voice pitch: a window into the communication of social power. *Current opinion in psychology*, 33:154–161, 2020.
- I. P. Banai, B. Banai, and K. Bovan. Vocal characteristics of presidential candidates can predict the outcome of actual elections. *Evolution and Human Behavior*, 38(3):309–314, 2017.
- N. M. Bauer. Shifting standards: How voters evaluate the qualifications of female and male candidates. *The Journal of Politics*, 82(1):1–12, 2020.
- N. M. Bauer and M. Santia. Going feminine: Identifying how and when female candidates emphasize feminine and masculine traits on the campaign trail. *Political Research Quarterly*, page 10659129211020257, 2021.
- R. Bernhard. The visual conjoint: A new solution for social desirability bias. *PolMeth Europe Conference Paper*, 2023.
- P. Boersma and D. Weenink. Praat: doing phonetics by computer (version 6.1.15), 2020.
- B. Borkowska and B. Pawlowski. Female voice frequency in the context of dominance and attractiveness perception. *Animal Behaviour*, 82(1):55–59, 2011.
- C. Boussalis, T. G. Coan, M. R. Holman, and S. Müller. Gender, candidate emotional expression, and voter reactions during televised debates. *American Political Science Review*, 115(4):1242–1257, 2021.
- C. Carpinella and N. M. Bauer. A visual analysis of gender stereotypes in campaign advertising. *Politics, Groups, and Identities*, 9(2):369–386, 2021.

- S. A. Collins. Men’s voices and women’s choices. *Animal behaviour*, 60(6):773–780, 2000.
- S. A. Collins and C. Missing. Vocal and visual attractiveness are related in women. *Animal behaviour*, 65(5):997–1004, 2003.
- C. Darwin. *The descent of man, and selection in relation to sex*. J. Murray, 1871.
- B. J. Dietrich, M. Hayes, and D. Z. O’Brien. Pitch Perfect: Vocal Pitch and the Emotional Intensity of Congressional Speech. *American Political Science Review*, 113(4):941–962, Nov. 2019.
- K. Dolan. The Impact of Gender Stereotyped Evaluations on Support for Women Candidates. *Political Behavior*, 32(1):69–88, Mar. 2010.
- K. Dolan. Gender Stereotypes, Candidate Evaluations, and Voting for Women Candidates: What Really Matters? *Political Research Quarterly*, 67(1):96–107, Mar. 2014.
- A. H. Eagly, A. B. Diekmann, M. C. Schneider, and P. Kulesa. Experimental Tests of an Attitudinal Theory of the Gender Gap in Voting. *Personality and Social Psychology Bulletin*, 29(10):1245–1258, Oct. 2003.
- V. S. Fahed, E. P. Doheny, M. Busse, J. Hoblyn, and M. M. Lowery. Comparison of acoustic voice features derived from mobile devices and studio microphone recordings. *Journal of Voice*, 2022.
- D. Feinberg, B. Jones, A. Little, D. Burt, and D. Perrett. Manipulations of fundamental and formant frequencies influence the attractiveness of human male voices. *Animal Behaviour*, 69(3):561–568, Mar. 2005a.
- D. R. Feinberg, B. C. Jones, L. M. DeBruine, F. R. Moore, M. J. Law Smith, R. E. Cornwell, B. P. Tiddeman, L. G. Boothroyd, and D. I. Perrett. The voice and face of woman: One ornament that signals quality? *Evolution and Human Behavior*, 26(5):398–408, Sept. 2005b.
- D. R. Feinberg, L. M. DeBruine, B. C. Jones, and D. I. Perrett. The Role of Femininity and Averageness of Voice Pitch in Aesthetic Judgments of Women’s Voices. *Perception*, 37(4):615–623, Apr. 2008.
- R. L. Fox and J. L. Lawless. Entering the arena? gender and the decision to run for office. *American Journal of Political Science*, 48(2):264–280, 2004.
- S. W. Gregory. Sounds of power and deference: Acoustic analysis of macro social constraints on micro interaction. *Sociological Perspectives*, 37(4):497–526, 1994.


- R. Johns and M. Shephard. Gender, candidate image and electoral preference. *The British Journal of Politics and International Relations*, 9(3):434–460, 2007.
- B. C. Jones, D. R. Feinberg, L. M. DeBruine, A. C. Little, and J. Vukovic. Integrating cues of social interest and voice pitch in men’s preferences for women’s voices. *Biology Letters*, 4(2):192–194, 2008.
- B. C. Jones, D. R. Feinberg, L. M. DeBruine, A. C. Little, and J. Vukovic. A domain-specific opposite-sex bias in human preferences for manipulated voice pitch. *Animal Behaviour*, 79(1):57–62, Jan. 2010.
- C. A. Klofstad. Candidate Voice Pitch Influences Election Outcomes: Voice Pitch Influences Voters. *Political Psychology*, 37(5):725–738, Oct. 2016.
- C. A. Klofstad. Looks and Sounds Like a Winner: Perceptions of Competence in Candidates’ Faces and Voices Influences Vote Choice. *Journal of Experimental Political Science*, 4(3): 229–240, 2017.
- C. A. Klofstad and R. C. Anderson. Voice pitch predicts electability, but does not signal leadership ability. *Evolution and Human Behavior*, 39(3):349–354, May 2018.
- C. A. Klofstad, R. C. Anderson, and S. Peters. Sounds like a winner: voice pitch influences perception of leadership capacity in both men and women. *Proceedings of the Royal Society B: Biological Sciences*, 279(1738):2698–2704, July 2012.
- C. A. Klofstad, R. C. Anderson, and S. Nowicki. Perceptions of Competence, Strength, and Age Influence Voters to Select Leaders with Lower-Pitched Voices. *PLOS ONE*, 10(8): e0133779, Aug. 2015.
- S. J. Ko, C. M. Judd, and D. A. Stapel. Stereotyping based on voice in the presence of individuating information: Vocal femininity affects perceived competence but not warmth. *Personality and Social Psychology Bulletin*, 35(2):198–211, 2009.
- J. W. Koch. Do Citizens Apply Gender Stereotypes to Infer Candidates’ Ideological Orientations? *The Journal of Politics*, 62(2):414–429, May 2000.
- D. E. Kroodsma. Using appropriate experimental designs for intended hypotheses in ‘song’playbacks, with examples for testing effects of song repertoire sizes. *Animal Behaviour*, 40(6):1138–1150, 1990.
- L. Laustsen, M. B. Petersen, and C. A. Klofstad. Vote Choice, Ideology, and Social Dominance Orientation Influence Preferences for Lower Pitched Voices in Political Candidates. *Evolutionary Psychology*, 13(3):147470491560057, Sept. 2015.

- L. Machlis, P. Dodd, and J. Fentress. The pooling fallacy: problems arising when individuals contribute more than one observation to the data set. *Zeitschrift für Tierpsychologie*, 68(3):201–214, 1985.
- C. H. Mo. The Consequences of Explicit and Implicit Gender Attitudes and Candidate Quality in the Calculations of Voters. *Political Behavior*, 37(2):357–395, 2015.
- F. Nagel, M. Maurer, and C. Reinemann. Is There a Visual Dominance in Political Communication? How Verbal, Visual, and Vocal Communication Shape Viewers’ Impressions of Political Candidates: Visual Dominance in Political Communication. *Journal of Communication*, 62(5):833–850, Oct. 2012.
- J. J. O’Connor and P. Barclay. The influence of voice pitch on perceptions of trustworthiness across social contexts. *Evolution and Human Behavior*, 38(4):506–512, July 2017.
- K. Pisanski, A. Groyecka-Bernard, and P. Sorokowski. Human voice pitch measures are robust across a variety of speech recordings: methodological and theoretical implications. *Biology letters*, 17(9):20210356, 2021.
- D. A. Puts, C. R. Hodges, R. A. Cárdenas, and S. J. Gaulin. Men’s voices as dominance signals: vocal fundamental and formant frequencies influence dominance attributions among men. *Evolution and Human Behavior*, 28(5):340–344, Sept. 2007.
- C. Rezlescu, T. Penton, V. Walsh, H. Tsujimura, S. K. Scott, and M. J. Banissy. Dominant Voices and Attractive Faces: The Contribution of Visual and Auditory Information to Integrated Person Impressions. *Journal of Nonverbal Behavior*, 39(4):355–370, Dec. 2015.
- C. Schild, E. Braunsdorf, K. Steffens, F. Pott, and J. Stern. Gender and context-specific effects of vocal dominance and trustworthiness on leadership decisions. *Adaptive Human Behavior and Physiology*, pages 1–19, 2022.
- M. C. Schneider and A. L. Bos. Measuring stereotypes of female politicians. *Political psychology*, 35(2):245–266, 2014.
- K. Searles, E. F. Fowler, T. N. Ridout, P. Strach, and K. Zuber. The Effects of Men’s and Women’s Voices in Political Advertising. *Journal of Political Marketing*, pages 1–29, May 2017.
- C. C. Tigue, D. J. Borak, J. J. O’Connor, C. Schandl, and D. R. Feinberg. Voice pitch influences voting behavior. *Evolution and Human Behavior*, 33(3):210–216, May 2012.
- I. R. Titze. Control of vocal intensity and efficiency. *Principles of voice production*, 1994.

- I. R. Titze. *Principles of Voice Production*. National Center for Voice and Speech, 2000.
- K. J. Tusing and J. P. Dillard. The sounds of dominance. vocal precursors of perceived dominance during interpersonal influence. *Human Communication Research*, 26(1):148–171, 2000.
- V. Uloza, N. Ulozaitė-Stanienė, T. Petrauskas, and R. Kregždytė. Accuracy of acoustic voice quality index captured with a smartphone—measurements with added ambient noise. *Journal of Voice*, 2021.
- B. Van der Woerd, M. Wu, V. Parsa, P. C. Doyle, and K. Fung. Evaluation of acoustic analyses of voice in nonoptimized conditions. *Journal of Speech, Language, and Hearing Research*, 63(12):3991–3999, 2020.
- R. H. Vieira, C. D. Gadenz, and M. Cassol. Estudo longitudinal de caracterização vocal em canto coral. *Revista CEFAC*, 17(6):1781–1791, Dec. 2015.
- S. E. Wolff and D. A. Puts. Vocal masculinity is a robust dominance signal in men. *Behavioral Ecology and Sociobiology*, 64(10):1673–1683, 2010.

A Online Appendix

A.1 Screenshot of the experiment

SABANCI
ÜNİVERSİTESİ

FACULTY OF
ARTS AND
SOCIAL SCIENCES

Lütfen seçeneklerdeki adayların **kişi başı sağlık harcamalarına ayıracakları yıllık miktarları** söyleyecekleri ses kayıtlarını dinleyin. Bu iki adaydan birisini seçmek zorunda olsaydınız oyunuzu hangi adaya verirdiniz?

▶ |

-0:01

»

☐

▶ |

-0:01

»

☐

→

Figure A.1: Screenshot of one choice task.

“Please listen to the audio recordings below where two candidates each declare how much annual budget they will allocate to per capita health expenditures. If you had to choose one of these two candidates, which candidate would you vote for?”

A.2 Descriptive Statistics and Balance Tests

Table 1: Summary Statistics

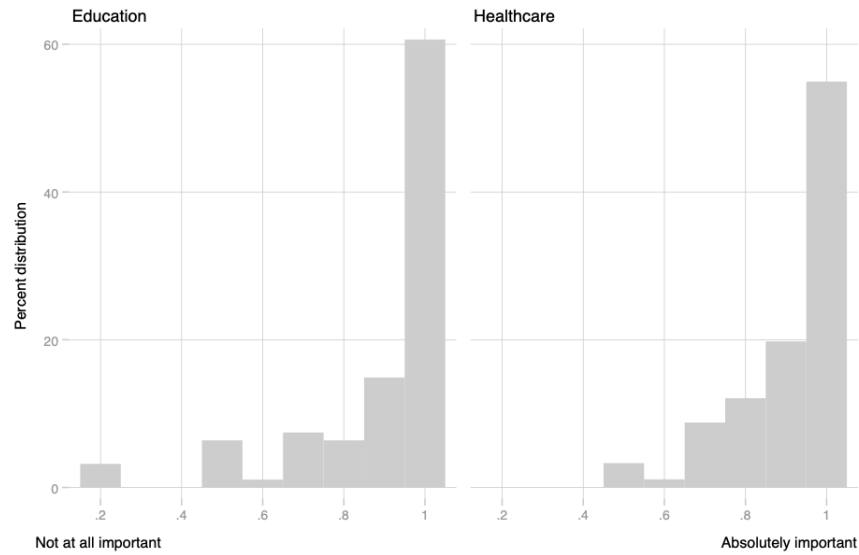
	Women Candidates Healthcare (N=44)	Women Candidates Education (N=46)	Men Candidates Healthcare (N=47)	Men Candidates Education (N=48)	Total Total (N=185)	p value
Gender						0.335
Female	18 (40.9%)	15 (32.6%)	16 (34.0%)	19 (39.6%)	68 (36.8%)	
Male	26 (59.1%)	31 (67.4%)	28 (59.6%)	28 (58.3%)	113 (61.1%)	
Other	0 (0.0%)	0 (0.0%)	3 (6.4%)	1 (2.1%)	4 (2.2%)	
Age						0.981
Mean (SD)	22.227 (1.939)	22.261 (1.639)	22.234 (1.507)	22.125 (1.632)	22.211 (1.669)	
Range	20 - 29	19 - 27	20 - 26	20 - 27	19 - 29	
Ideology						0.774
height Mean (SD)	0.455 (0.504)	0.370 (0.488)	0.404 (0.496)	0.354 (0.483)	0.395 (0.490)	
Range	0 - 1	0 - 1	0 - 1	0 - 1	0 - 1	
Turnout						0.850
Mean (SD)	0.955 (0.211)	0.913 (0.285)	0.915 (0.282)	0.938 (0.245)	0.930 (0.256)	
Range	0 - 1	0 - 1	0 - 1	0 - 1	0 - 1	
Trust						0.831
Most people can be trusted	43 (97.7%)	44 (95.7%)	45 (95.7%)	45 (93.8%)	177 (95.7%)	
People should be approached carefully	1 (2.3%)	2 (4.3%)	2 (4.3%)	3 (6.2%)	8 (4.3%)	
Satisfaction from education/healthcare						0.001
Mean (SD)	3.409 (0.996)	1.913 (1.029)	3.511 (1.040)	1.750 (0.911)	2.632 (1.283)	
Range	1 - 5	1 - 5	1 - 5	1 - 5	1 - 5	
Importance of public service provision (education/healthcare)						0.775
Mean (SD)	0.898 (0.147)	0.887 (0.183)	0.919 (0.110)	0.890 (0.201)	0.898 (0.164)	
Range	0.500 - 1.000	0.200 - 1.000	0.600 - 1.000	0.200 - 1.000	0.200 - 1.000	
“In general, do you think female politicians or male politicians are more successful in addressing education/ healthcare-related issues?”						0.086
Men	3 (6.8%)	10 (21.7%)	4 (8.5%)	3 (6.2%)	20 (10.8%)	
Women	3 (6.8%)	8 (17.4%)	6 (12.8%)	7 (14.6%)	24 (13.0%)	
No difference	38 (86.4%)	28 (60.9%)	37 (78.7%)	38 (79.2%)	141 (76.2%)	

Table 2: Balance table

	Treatment -Candidate Gender (Woman=0, Man=1)
(Intercept)	0.851*
	(0.424)
Gender	-0.045
(Man=1, Woman=0)	(0.078)
Income	-0.001
	(0.011)
Turnout	-0.052
(voted in the last election=1)	(0.156)
Ideology	-0.218
	(0.192)
Trust	-0.093
	(0.197)
R ²	0.013
Adj. R ²	-0.026
Num. obs.	185

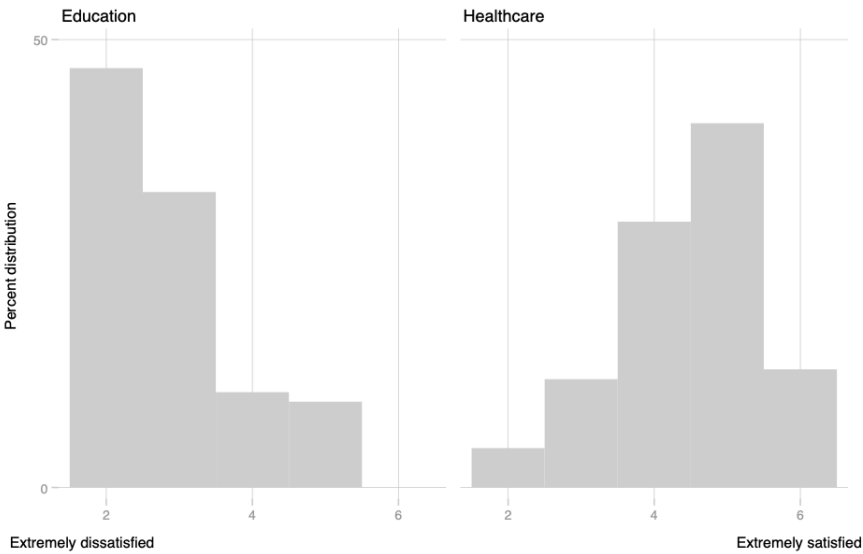
*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Figure A.2: Importance of government spending for education and health



Survey Question: “How important do you think it is that governments provide education/health services?” ($N = 185$)

Figure A.3: Satisfaction from Education and Health Services



Survey Question: “How satisfied are you with the education/healthcare system in Turkey?” ($N = 185$)

A.3 Validation of the perceived voice pitch difference

In order to verify whether the alteration of vocal pitch within a single individual yielded an audible distinction between the two modified recordings and to ascertain the possibility of the modified voices being attributed to separate individuals, a separate sample of participants ($n = 95$) was used. First, 87.37% of our sample states the likelihood of the recordings belonging to two distinct individuals to be positive, and hence that there is a noticeable difference between the two recordings ($t = 62.734$, $p = 0.001$). Second, we also analyzed if the participants who listened to men’s voices were different from those who listened to women’s voices in terms of the perceived difference between the higher and the lower pitch versions of the same recording. It turns out that there is no significant difference between those who listened to men versus women ($t = 1.1803$, $p = 0.2388$). Thus, the Praat voice pitch manipulations were equally successful in generating higher and lower pitch voices for both man and woman candidates.

A.4 Regression Tables

Table 3: Linear Regression of the Probability of Voting for the LP candidate on Candidate Gender

	0 TL	0 TL	200 TL	200 TL	400 TL	400 TL	600 TL	600 TL	800 TL	800 TL	1000 TL	1000 TL
(Intercept)	0.60*** (0.03)	0.41 (0.28)	0.64*** (0.04)	0.69* (0.36)	0.61*** (0.03)	1.10*** (0.33)	0.59*** (0.04)	0.47 (0.37)	0.59*** (0.04)	0.45 (0.38)	0.48*** (0.04)	0.74 (0.48)
Candidate Gender (Woman=0, Man=1)	0.09** (0.05)	0.07 (0.05)	-0.12** (0.05)	-0.14*** (0.05)	-0.08 (0.05)	-0.11** (0.05)	-0.06 (0.06)	-0.10* (0.06)	-0.13** (0.05)	-0.15*** (0.06)	-0.04 (0.05)	-0.06 (0.05)
Covariate adjusted	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
R ²	0.02	0.12	0.03	0.16	0.01	0.13	0.01	0.13	0.03	0.12	0.00	0.13
Adj. R ²	0.02	0.03	0.02	0.07	0.01	0.04	0.00	0.04	0.02	0.03	-0.00	0.03
Num. obs.	185	185	185	185	185	185	185	185	185	185	185	185

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$. Note: Covariates are gender, age, income, ideology, past turnout, general trust of the respondent.

Table 4: Linear Regression of the Probability of Voting for the LP candidate on Candidate Gender (excluding respondents who experienced some trouble listening to the experimental stimuli (n=4))

	0 TL	0 TL	200 TL	200 TL	400 TL	400 TL	600 TL	600 TL	800 TL	800 TL	1000 TL	1000 TL
(Intercept)	0.60*** (0.03)	0.59** (0.26)	0.64*** (0.04)	0.74* (0.38)	0.62*** (0.03)	1.11*** (0.34)	0.60*** (0.04)	0.57 (0.39)	0.59*** (0.04)	0.46 (0.41)	0.48*** (0.04)	0.89* (0.50)
Candidate Gender (Woman=0, Man=1)	0.10** (0.04)	0.07 (0.05)	-0.13** (0.05)	-0.15*** (0.05)	-0.09 (0.05)	-0.11** (0.05)	-0.06 (0.06)	-0.10* (0.06)	-0.13** (0.05)	-0.16*** (0.06)	-0.04 (0.05)	-0.06 (0.05)
Covariate adjusted	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
R ²	0.03	0.12	0.03	0.15	0.01	0.12	0.01	0.12	0.03	0.12	0.00	0.13
Adj. R ²	0.02	0.02	0.03	0.06	0.01	0.02	0.00	0.03	0.03	0.02	-0.00	0.03
Num. obs.	181	181	181	181	181	181	181	181	181	181	181	181

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

Table 5: Linear Regression of the Probability of Voting for the LP candidate on Candidate Gender (covariate adjustment also included completion time, listening medium, and listening device)

	0 TL	200 TL	400 TL	600 TL	800 TL	1000 TL
(Intercept)	0.41 (0.28)	0.76** (0.36)	1.13*** (0.35)	0.55 (0.36)	0.50 (0.41)	0.67 (0.47)
Candidate Gender (Woman=0, Man=1)	0.08 (0.05)	-0.14*** (0.05)	-0.09* (0.05)	-0.09 (0.06)	-0.14** (0.06)	-0.06 (0.06)
Covariate adjusted	Yes	Yes	Yes	Yes	Yes	Yes
R ²	0.13	0.18	0.14	0.16	0.14	0.16
Adj. R ²	0.02	0.07	0.02	0.05	0.02	0.05
Num. obs.	179	179	179	179	179	179

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.