**Online Appendix for *JEPS* Article “Looks and Sounds Like a Winner” (Klofstad)**

*Demographic Survey Questions*

Age

“What year were you born?”

Income

“Before taxes and other deductions, what is your annual household income?”

-Under $15,000

-Between $15,000 and $29,000

-Between $30,000 and $49,999?

-Between $50,000 and $74,999

-Between $75,000 and $99,999

-Between $100,000 and $199,999

-More than $199,999

Partisanship

“Generally speaking, do you consider yourself a Democrat, a Republican, an Independent, or something else?”

-Democrat

-Republican

-Independent

-Other

-None

-I don't know

“Would you call yourself a strong Democrat or a not very strong Democrat?”

-Strong

-Not strong

“Would you call yourself a strong Republican or a not very strong Republican?”

-Strong

-Not strong

“Do you think of yourself as closer to the Republican or Democratic Party?”

-Republican

-Democratic

-Neither

Ideology

“Where would you place yourself on a scale that goes from ‘extremely liberal’ to ‘extremely conservative,’ or haven’t you thought much about this?”

-Extremely liberal

-Liberal

-Slightly liberal

-Moderate or middle of the road

-Slightly conservative

-Conservative

-Extremely conservative

-I haven't thought much about this

Interest in Politics

“Some people seem to follow what's going on in government and public affairs most of the time, whether there's an election going on or not. Others aren't that interested. Would you say you follow what's going on in government and public affairs:”

-Most of the time

-Some of the time

-Only now and then

-Hardly at all

Education

“What is the highest level of education you have completed?”

-Less than high school

-High school graduate or GED

-Some college, but no degree (yet)

-2-year college degree

-4-year college degree

-Post-graduate degree (MA, MBA, MD, JD, PhD, etc.)

*Descriptive Statistics*

|  |
| --- |
| Table A1. Descriptive statistics on simulated elections experiment subjects |
|  | Min | Max | Mean | SD |
| Age (in years) | 18 | 93 | 46.49 | 17.15 |
| Income ($24,999 or less - $100,000 or more) | 1 | 4 | 2.43 | 1.07 |
| Partisanship (Dem - Rep) | 1 | 7 | 3.67 | 2.06 |
| Ideology (Lib - Con) | 1 | 7 | 4.04 | 1.75 |
| Interest in Politics (high - low) | 1 | 4 | 3.03 | .95 |
| Education (less than high school - post-grad degree) | 1 | 6 | 3.78 | 1.39 |

|  |
| --- |
| Table A2. Descriptive statistics on variables included in linear regression analysis of simulated elections experiment (Table 1) |
|  | Min | Max | Mean | SD |
| Candidate has competent face | 0 | 1 | .5 | .50 |
| Candidate has competent voice | 0 | 1 | .5 | .50 |
| Competent face\*Competent voice | 0 | 1 | .25 | .43 |
| Candidate is female | 0 | 1 | .5 | .50 |
| Competent face\*female | 0 | 1 | .25 | .43 |
| Competent voice\*female | 0 | 1 | .25 | .43 |
| Competent face\*Competent voice\*Female | 0 | 1 | .13 | .33 |

|  |
| --- |
| Table A3. Descriptive statistics on simulated elections experiment subjects by treatment block |
|  | Block 1 | Block 2 | Block 3 | Block 4 |
| Age (in years) | 46.49(17.15) | 46.49(17.15) | 46.49(17.15) | 46.49(17.15) |
| Income ($24,999 or less - $100,000 or more) | 2.43(1.07) | 2.43(1.07) | 2.43(1.07) | 2.43(1.07) |
| Partisanship (Dem - Rep) | 3.67(2.06) | 3.67(2.06) | 3.67(2.06) | 3.67(2.06) |
| Ideology (Lib - Con) | 4.04(1.75) | 4.04(1.75) | 4.04(1.75) | 4.04(1.75) |
| Interest in Politics (high - low) | 3.03(.95) | 3.03(.95) | 3.03(.95) | 3.03(.95) |
| Education (less than high school - post-grad degree) | 3.78(1.39) | 3.78(1.39) | 3.78(1.39) | 3.78(1.39) |
| Sex (female) | .50(.50) | .50(.50) | .50(.50) | .50(.50) |
| Note: Cell entries are means with standard deviations in parentheses. Values do not vary because each subject participated in elections in each of the four blocks. More specifically, each subject participated in six randomly selected elections from each block, for a total of 24 elections. Descriptions of the experimental blocks are detailed in the Election Experiment Procedures section below, and in the main manuscript. |

*Facial Stimuli Procedures*

A survey was administered to 394 subjects (197 men and 197 women) aged 18 years or older by Qualtrics, an online service used to develop and administer questionnaires on the Internet. The study was fielded between 01/30/2015 and 03/03/2015. Qualtrics partnered with Clear Voice Surveys (i.e., Clear Voice) to recruit a national sample of subjects that matched U.S. Census records on age and income. Based on this quota-based recruitment procedure, there is no response or completion rate to report, nor was there any subject attrition. Clear Voice maintains panels of subjects that are only used for research. Individuals voluntarily join a Clear Voice panel through their website, or by responding to an online advertisement (e.g., a banner advertisement on a website). Clear Voice complies fully with European Society for Opinion and Marketing Research (ESOMAR) standards for protecting research subjects’ privacy and information. Subjects received $.50 USD in exchange for voluntary participation in the study. They were invited to participate by email, and consented voluntarily to particulate by clicking a link to the survey in that email. Subjects were free to end participation at any time by closing their web browser. Approval to conduct research with human subjects was granted by the University of Miami Human Subject Research Office on January 23, 2015 (Protocol #20150071).

 Pictures of the faces of Members of the 113th House of Representatives from the 2013 Congressional Pictorial Directory (U.S. Government Printing Office 2013) were used as facial stimuli. Each picture in the Directory is standardized so the Member appears from the shoulders up in front of a neutral background. To further standardize the pictures they were converted from color to grayscale. A random sample of 25 female (12 Democrats and 13 Republicans) and 25 male (11 Democrats and 14 Republicans) Member’s pictures was selected from the Directory to present to subjects. In line with Rule and Ambady (2010), Members from racial or ethnic minority groups were excluded from the random sample of pictures to avoid adding racial bias as a factor in the study. Prominent Members, such as those who had recently run for national office or who held leadership positions in the House (e.g., Paul Ryan, Michelle Bachman, Nancy Pelosi, John Boehner, and the like), were also excluded to reduce the potential for partisan bias in the study (e.g., subjects who identify with the Democratic Party might have a stronger affinity for recognizable Democratic politicians).

 After completing a brief demographic questionnaire, subjects were asked to rate the competence of each of the 50 faces. The faces were grouped into two blocks by sex of Member. The order in which these blocks were presented to the subjects, as well as the order of the faces within each block, was randomized. For each face subjects were asked, “On a scale of 0-10, where 0 means ‘very incompetent’ and 10 means ‘very competent,’ how would you rate the person in the picture?” The questionnaire was designed to take no more than 15 minutes to complete.

*Vocal Stimuli Procedures*

Five male and five female native English speakers were recorded saying the sentence, “I urge you to vote for me this November,” a politically relevant yet partisan neutral statement. The women ranged in age from 21 to 38 years (x̄ = 26 years, SE = 3), and the men 20 to 41 years (x̄ = 28 years, SE = 4). Voices were recorded as .wav files in an Acoustic Systems soundproof room using a Shure SM57 microphone and a Marantz PMD660 solid-state recorder. Each audio file was inspected aurally and visually in Audacity (v. 2.0.1; audacity.sourceforge.net) to ensure that they were free from speech errors and non-speech noise. Engineering Design’s Signal acoustics analysis program (v. 4.02.04; www.engdes.com) was used to normalize the amplitude (i.e., “loudness”) of the recordings. The pitch of female and male voices ranged from 189 to 207 Hz (x̄ = 199 Hz, SE = 3 Hz) and 91 to 116 Hz (x̄ = 107 Hz, SE = 4 Hz) respectively. Praat (v. 5.1.43; Boersma & Weenink 2013) was used to measure the pitch of the recordings.

 The vocal stimuli used in the experiment are pairs of digitally manipulated versions of these ten voice recordings. In each pair, one version was manipulated to be higher than the original recording, and the other to be lower. Following previous studies (Jones et al. 2008) the pitch of each original recording was altered +/-.5 equivalent rectangular bandwidths (ERB) with Praat (v. 5.1.43; Boersma & Weenink 2013), which uses the Pitch Synchronous Overlap Add Method (PSOLA) algorithm to alter F0 (Jones et al. 2008). The relationship between absolute and perceived pitch in humans is logarithmic. Manipulation by ERB accounts for this nonlinearity, and produces a constant perceivable gap between the higher- and lower-pitched sound files regardless of the pitch of the original recording. The magnitude of a +/-.5 ERB shift is equivalent to a gap of 40 Hz between the higher and lower version of each pair of altered recordings, which is equivalent to difference between open string G and D notes on a guitar, and the difference between C4 (“Middle C”) and A3 on a piano. The pitch of the higher- and lower-pitched female sound files ranged from 214 to 233 Hz (x̄ = 224 Hz, SE = 3 Hz) and 170 to 190 Hz (x̄ = 181 Hz, SE = 3 Hz) respectively. The pitch of the higher- and lower-pitched male sound files ranged from 110 to 136 Hz (x̄ = 127 Hz, SE = 4 Hz) and 81 to 98 Hz (x̄ = 91 Hz, SE = 3 Hz) respectively. Praat (v. 5.1.43; Boersma & Weenink 2013) was used to measure the pitch of the recordings.

A previous study using these same vocal stimuli verified that voters can perceive which voice of each pair is higher in pitch (Klofstad et al. 2012), and other studies using these same stimuli have shown that the lower voices of each pair are perceived as more competent (Klofstad et al. 2012) and are more likely to win elections (Anderson & Klofstad 2012; Klofstad 2016; Klofstad et al. 2012).

*Election Experiment Procedures*

The election experiment was administered online to 840 subjects (420 men and 420 women) aged 18 years or older by Qualtrics. The study was fielded between 04/15/2015 and 04/24/2015. Qualtrics partnered with Clear Voice Surveys (i.e., Clear Voice) to recruit a national sample of subjects that matched U.S. Census records on age and income. Based on this quota-based recruitment procedure, there is no response or completion rate to report, nor was there any subject attrition. Clear Voice maintains panels of subjects that are only used for research. Individuals voluntarily join a Clear Voice panel through their website, or by responding to an online advertisement (e.g., a banner advertisement on a website). Clear Voice complies fully with European Society for Opinion and Marketing Research (ESOMAR) standards for protecting research subjects’ privacy and information. Subjects received $.50 USD in exchange for voluntary participation in the study. They were invited to participate by email, and consented voluntarily to particulate by clicking a link to the survey in that email. Subjects were free to end participation at any time by closing their web browser. Approval to conduct research with human subjects was granted by the University of Miami Human Subject Research Office on January 23, 2015 (Protocol #20150071).

Before participating in the experiment subjects answered a brief demographic questionnaire and completed a sound check task to ensure that they could hear audio played by the online survey instrument. More specifically, subjects listened to a recording that contained the correct response to a survey question. If the correct response was not provided the subject should not continue participating in the study. Subjects chose whether to use computer speakers (N = 703) or headphones (N = 137). Experiments conducted online that use vocal stimuli where voice pitch has been manipulated digitally produce results that are comparable to experiments conducted in a laboratory setting (Feinberg et al. 2008).

In each simulated election subjects were instructed to, “Please look at the photos and use the audio players to hear each candidate's voice.” They were then asked, “If these two candidates were running against each other in an election, who would you vote for?” A specific type of leadership role was not referred to in the question to remove this potentially confounding factor from the experiment (e.g., Anderson & Klofstad 2012). The elections were sorted into four blocks:

* Block 1: male competent face paired with competent voice vs. male incompetent face paired with incompetent voice (N = 50 elections)
* Block 2: male competent face paired with incompetent voice vs. male incompetent face paired with competent voice (N = 50 elections)
* Block 3: female competent face paired with competent voice vs. female incompetent face paired with incompetent voice (N = 50 elections)
* Block 4: female competent face paired with incompetent voice vs. female incompetent face paired with competent voice (N = 50 elections)

To make the task less burdensome each subject participated in six randomly selected elections from each block, for a total of 24 elections. The order in which the four blocks of elections were presented, and the order in which the elections within each block were presented, was randomized. Subjects participated in multiple simulated elections with different candidates to reduce biases due to pseudoreplication, whereby the idiosyncratic characteristics of any one pair of candidates might influence the results of the experiment (Kroodsma 1990; Machlis et al. 1985). Each election was participated in by an average of 84 subjects (minimum = 59, maximum = 112, SE = .44). The questionnaire was designed to take no more than 15 minutes to complete.

*Supplementary Data Analyses*

“Outlier” Faces

Some facial stimuli may have been perceived significantly more or less favorably than others. This was tested by conducting a one way between subjects ANOVA with candidate face as the factor for the following groups: (1) males with competent faces (F4,100 = .68, p = .61), (2) males with incompetent faces (F4,100 = 8.09, p < .001), (3) females with competent faces (F4,100 = 3.70, p =.01), and (4) females with incompetent faces (F4,100 = 1.47, p = .22). In the case of males with incompetent faces there were two faces that earned significantly fewer votes compared to two other faces. In the case of females with competent faces there was one face that performed significantly better than one other face. The analyses presented in the paper were replicated after excluding these under- and over-performing cases from the data set. The bivariate results remain unchanged (faces: t340 = -20.33, p < .001; voices: t340 = -5.54, p < .001). As documented below in Table A4, the multivariate results are largely unchanged, with the exception of the insignificant result for *Candidate has competent voice* (p = .297) in the full-form model. This, however, is in line with the main finding of the paper that the visual signal mattered more to voters than the vocal signal.

|  |
| --- |
| Table A4. Influence of candidate facial and vocal competence on vote share in simulated elections |
|  | Original | Outlier faces kicked out | Original | Outlier faces kicked out |
| Candidate has competent face | 23.63\*\*\*(1.28) | 22.44\*\*\*(1.40) | 22.49\*\*\*(1.68) | 18.85\*\*\*(1.73) |
| Candidate has competent voice | 8.56\*\*\*(1.28) | 9.06\*\*\*(1.40) | 3.33\*(1.68) | 1.87(1.79) |
| Competent face\*Competent voice | -.003(.002) | -1.06(.85) | < .001(< .001) | 1.46(1.35) |
| Candidate is female | --- | --- | -6.37\*\*\*(1.74) | -10.01\*\*\*(1.79) |
| Competent face\*Female | --- | --- | 2.28(2.46) | 4.36(2.61) |
| Competent voice\*Female | --- | --- | 10.47\*\*\*(2.46) | 11.93\*\*\*(2.54) |
| Competent face\*Competent voice\*Female | --- | --- | -.006(.003) | -1.11(1.69) |
| Constant | 33.91\*\*\*(.92) | 34.70\*\*\*(1.03) | 37.09\*\*\*(1.30) | 40.73(1.37) |
|  |  |  |  |  |
| R2 | .66 | .63 | .69 | .68 |
| N | 400 | 342 | 400 | 342 |
| \*p ≤ .05; \*\*\*p ≤ .001 (robust standard errors, clustered by election, in parentheses)Note: Cell entries are linear regression coefficients.  |

*Multilevel Regression Model*

A reviewer questioned whether a multilevel regression model approach might be more appropriate than clustering standard errors by election. If a multilevel approach (i.e., random intercepts by election) is used (Table A5), the only change is that the *Candidate has competent voice coefficient* in the full-form model is no longer significant (p = .053). The said the coefficient is very close to p < .05, and this weakening of the voice result is in line with the main finding of the paper that the visual signal mattered more to voters than the vocal signal.

|  |
| --- |
| Table A5. Influence of candidate facial and vocal competence on vote share in simulated elections |
|  | Original | Multilevel model | Original | Multilevel model |
| Candidate has competent face | 23.63\*\*\*(1.28) | 23.63\*\*\*(1.28) | 22.49\*\*\*(1.68) | 22.49\*\*\*(1.72) |
| Candidate has competent voice | 8.56\*\*\*(1.28) | 8.56\*\*\*(1.28) | 3.33\*(1.68) | 3.33^(1.72) |
| Competent face\*Competent voice | -.003(.002) | -.003(1.80) | < .001(< .001) | < .001(2.43) |
| Candidate is female | --- | --- | -6.37\*\*\*(1.74) | -6.37\*\*\*(1.72) |
| Competent face\*Female | --- | --- | 2.28(2.46) | 2.28(2.43) |
| Competent voice\*Female | --- | --- | 10.47\*\*\*(2.46) | 10.47\*\*\*(2.43) |
| Competent face\*Competent voice\*Female | --- | --- | -.006(.003) | -.006(3.44) |
| Constant | 33.91\*\*\*(.92) | 33.91\*\*\*(.90) | 37.09\*\*\*(1.30) | 37.09\*\*\*(3.44) |
| Election random effects parameter | --- | < .001(<.001) | --- | < .001(<.001) |
|  |  |  |  |  |
| R2 | .66 | --- | .69 | --- |
| Log-likelihood | --- | -1447.17 | --- | -1428.69 |
| N | 400 | 400 | 400 | 400 |
| ^p < .10; \*p ≤ .05; \*\*\*p ≤ .001 (robust standard errors, clustered by election, in parentheses for “Original” models)Note: Cell entries are linear regression coefficients.  |

*Online Appendix References*

Anderson, RC, & Klofstad, CA 2012 Preference for leaders with masculine voices

holds in the case of feminine leadership roles. *PLoS ONE* 7, e51216.

Boersma P, & Weenink D 2013 Praat: doing phonetics by computer. www.praat.org

Feinberg DR, DeBruine LM, Jones BC, & Perrett DI 2008 The role of femininity and

averageness of voice pitch in aesthetic judgments of women's voices. *Perception* 37, 615-623.

Klofstad, CA 2016 Candidate voice pitch influences election outcomes. *Polit. Psychol*.

37, 725-738.

Klofstad, CA, Anderson, RC, & Peters, S 2012 Sounds like a winner: voice pitch

influences perception of leadership capacity in both men and women. *P. R. Soc. B.* 297, 2698-2704.

Kroodsma DE 1990 Using appropriate experimental-designs for intended hypotheses in

song playbacks, with examples for testing effects of song repertoire sizes. *Anim. Behav.* 40, 1138-1150.

Machlis L, Dodd PWD, & Fentress JC 1985 The pooling fallacy—problems arising when

individuals contribute more than one observation to the data set. *Z. Tierpsychol* 68, 201–214.

Rule, NO, Ambady, N, Adams, RGB, Jr., et al. 2010 Polling the face: prediction and

consensus across cultures. *J. Pers. Soc. Psychol.* 98, 1-15.