Supplementary Material

Comparing alignment toward American, British, and Indian English text-to-speech (TTS) voices: Influence of social attitudes and talker guise

Nicole Dodd\*, Michelle Cohn, Georgia Zellou

**\* Correspondence:** Nicole Dodd: ncdodd@ucdavis.edu

# Supplement A: TTS Voice Norming Study

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Dialect** | **Voice** | **Accent strength (mean)** | **Human-likeness (mean)** | **Age  (mean)** |
| US | AWS Polly Joanna (neural) | 66.85 | 38.49 | 35.75 |
| AWS Polly Kendra (neural) | 61.87 | 42.49 | 42.3 |
| AWS Polly Kimberly (neural) | 52.4 | 24.22 | 34.55 |
| AWS Polly Salli (neural) | 65.84 | 62.47 | 26.7 |
| Google TTS US1 | 61.61 | 23.33 | 41.27 |
| Google TTS US2 | 62.18 | 20.96 | 35.95 |
| Apple Siri Allison | 50.64 | 17.4 | 37.998 |
| Apple Siri Ava | 51.67 | 22.81 | 41.65 |
| Apple Siri Susan | 55.95 | 20.85 | 34.9 |
| RP | AWS Polly Amy | 61.04 | 68.71 | 34.53 |
| AWS Polly Emma | 58.7 | 63.96 | 23.71 |
| Google TTS GB1 | 57.38 | 42.45 | 26.49 |
| Google TTS GB2 | 69.51 | 52.02 | 31.69 |
| Apple Siri Serena | 66.04 | 47.71 | 45.18 |
| Apple Siri Kate | 66.42 | 31.78 | 42.85 |
| IN | AWS Polly Aditi | 58.18 | 57.29 | 30.64 |
| AWS Polly Raveena | 55.85 | 39.53 | 39.31 |
| Google TTS IN1 | 71.75 | 64.07 | 23.95 |
| Google TTS IN2 | 51.15 | 46.58 | 39.12 |
| Apple Siri Veena | 50.64 | 47.42 | 29.2 |

**Table S1.** Ratings for all voices used in the voice norming study. We selected all available female TTS voices for each dialect in Amazon Polly, Google TTS, and Apple Siri voices. At the time of the study, Amazon offered neural options for US voices, but only concatenative TTS for RP or IN voices.

# Supplement B: Interactive Task

|  |  |
| --- | --- |
| **Conversation piece** | **Script** |
| Opening remarks | Hey, my name is Sarah. Let’s get started. [AWS Salli neural]  Hi, my name is Elizabeth. Let’s jump right in. [AWS Joanna neural]  Hi, I’m Rebecca. Let’s move right along. [AWS Amy]  Hello, my name is Charlotte. Alright, ready to go. [Google GB2]  Hey, my name is Shreya. Ready to go. [AWS Aditi]  Hi, my name is Meera. Let’s get started. [Google IN1] |
| Acknowledgement of answer  (one after each question; randomly distributed) | Great. | Okay. | Cool. | Got it. | Alrighty. | Awesome. | Okay, thanks. |
| Closing remarks | That’s all, thanks. | All done, thanks. | That’s all for now, thanks. |

**Table S2.** A comprehensive list of all conversational snippets used in the interactive task. Each voice had a distinct opening remark with a culturally appropriate name. Acknowledgments of answers and closing remarks were randomly assigned.

# Supplement C: Full Model Estimates

|  |  |  |  |
| --- | --- | --- | --- |
| *Composite* | *Coef* | *SE* | *p* |
| **(Intercept)** | **64.07** | **0.83** | **<0.001** |
| **Dialect (IN)** | **3.96** | **1.18** | **<0.001** |
| **Dialect (RP)** | **3.88** | **1.18** | **0.001** |
| **Dialect (US)** | **-7.84** | **1.18** | **<0.001** |
| **Guise (Device)** | **2.09** | **0.83** | **0.01** |
| **Guise (Human)** | **-2.09** | **0.83** | **0.01** |
| IN:device | -0.66 | 1.18 | 0.57 |
| RP:device | -0.76 | 1.18 | 0.52 |

**Table S3**. Full model estimates for composite attitude scores linear regression, including re-leveled fixed effects. Significant effects are bolded.

|  |  |  |  |
| --- | --- | --- | --- |
| *Friendliness* | *Coef* | *SE* | *p* |
| **(Intercept)** | **63.03** | **1.24** | **<0.001** |
| **Dialect (IN)** | **9.16** | **1.76** | **<0.001** |
| Dialect (RP) | 3.17 | 1.76 | 0.07 |
| **Dialect (US)** | **-12.34** | **1.76** | **<0.001** |
| Guise (Device) | 1.17 | 1.24 | 0.35 |
| Guise (Human) | -1.17 | 1.24 | 0.35 |
| IN:device | 0.87 | 1.76 | 0.62 |
| RP:device | -1.38 | 1.76 | 0.43 |

**Table S4**. Full model estimates for friendliness scores linear regression, including re-leveled fixed effects. Significant effects are bolded.

|  |  |  |  |
| --- | --- | --- | --- |
| *Intelligence* | *Coef* | *SE* | *p* |
| **(Intercept)** | **65.30** | **1.08** | **<0.001** |
| **Dialect (IN)** | **3.49** | **1.53** | **0.02** |
| **Dialect (RP)** | **3.57** | **1.53** | **0.02** |
| **Dialect (US)** | **-7.06** | **1.53** | **<0.001** |
| **Guise (Device)** | **2.93** | **1.08** | **0.01** |
| **Guise (Human)** | **-2.93** | **1.08** | **0.01** |
| IN:device | -0.79 | 1.53 | 0.61 |
| RP:device | -0.58 | 1.53 | 0.70 |

**Table S5**. Full model estimates for intelligence scores linear regression, including re-leveled fixed effects. Significant effects are bolded.

|  |  |  |  |
| --- | --- | --- | --- |
| *Socioeconomic* | *Coef* | *SE* | *p* |
| **(Intercept)** | **63.87** | **0.92** | **<0.001** |
| Dialect (IN) | -0.78 | 1.30 | 0.55 |
| **Dialect (RP)** | **4.89** | **1.30** | **<0.001** |
| **Dialect (US)** | **-4.11** | **1.30** | **0.002** |
| **Guise (Device)** | **2.18** | **0.92** | **0.02** |
| **Guise (Human)** | **-2.18** | **0.92** | **0.02** |
| IN:device | -2.07 | 1.30 | 0.11 |
| RP:device | -0.33 | 1.30 | 0.80 |

**Table S6**. Full model estimates for perceived socioeconomic status scores linear regression, including re-leveled fixed effects. Significant effects are bolded.

|  |  |  |  |
| --- | --- | --- | --- |
| *Naturalness* | *Coef* | *SE* | *p* |
| **(Intercept)** | **38.19** | **1.47** | **<0.001** |
| **Dialect (IN)** | **13.23** | **2.07** | **<0.001** |
| Dialect (RP) | 1.58 | 2.07 | 0.45 |
| **Dialect (US)** | **-14.81** | **2.07** | **<0.001** |
| **Guise (Device)** | **3.42** | **1.47** | **0.02** |
| **Guise (Human)** | **-3.42** | **1.47** | **0.02** |
| IN:device | 0.46 | 2.07 | 0.82 |
| RP:device | -1.73 | 2.07 | 0.40 |

**Table S7**. Full model estimates for naturalness scores linear regression, including re-leveled fixed effects. Significant effects are bolded.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *Coef* | *SE* | *Lower 95% CI* | *Upper 95% CI* |
| (Intercept) | 0.03 | 0.03 | -0.04 | 0.09 |
| **Dialect (IN)** | **0.08** | **0.03** | **0.03** | **0.14** |
| **Dialect (US)** | **-0.09** | **0.03** | **-0.14** | **-0.03** |
| **Guise (Device)** | **-0.05** | **0.03** | **-0.11** | **0.01** |
| Attitude | 0.0 | 0.03 | -0.05 | 0.06 |
| Naturalness | -0.01 | 0.03 | -0.07 | 0.04 |
| IN:device | 0.04 | 0.03 | -0.02 | 0.09 |
| US:device | 0.01 | 0.03 | -0.05 | 0.07 |
| IN:attitude | -0.06 | 0.04 | -0.15 | 0.02 |
| US:attitude | 0.03 | 0.04 | -0.04 | 0.11 |
| device:attitude | 0.02 | 0.03 | -0.03 | 0.08 |
| IN:naturalness | 0.06 | 0.04 | -0.02 | 0.13 |
| US:naturalness | -0.01 | 0.04 | -0.08 | 0.07 |
| device:naturalness | -0.03 | 0.03 | -0.08 | 0.02 |
| attitude:naturalness | 0.03 | 0.03 | -0.03 | 0.09 |
| IN:device:attitude | -0.03 | 0.04 | -0.11 | 0.05 |
| US:device:attitude | -0.01 | 0.04 | -0.08 | 0.07 |
| IN:device:naturalness | 0.02 | 0.04 | -0.06 | 0.10 |
| US:device:naturalness | -0.01 | 0.04 | -0.08 | 0.07 |
| IN:attitude:naturalness | 0.01 | 0.05 | -0.09 | 0.11 |
| US:attitude:naturalness | -0.01 | 0.04 | -0.09 | 0.07 |
| device:attitude:naturalness | 0.01 | 0.03 | -0.06 | 0.07 |
| IN:device:attitude:naturalness | 0.00 | 0.05 | -0.09 | 0.10 |
| US:device:attitude:naturalness | -0.01 | 0.04 | -0.09 | 0.06 |

**Table S8.** Full model estimates for AXB ratings logistic regression. Reliable effects are bolded.