The Effects of Voice Pitch on Perceptions of Robots

Christopher Peterson, University of Nevada, Reno
Andrew Palmer, David Feil-Seifer
chrispeterson@nevada.unr.edu

Introduction

Prior research in human-robot interaction (HRI) has shown people respond similarly to vocal cues in synthetic robot speech as they do in human speech. However, the effects of voice pitch have been minimally researched. This study investigates whether voice pitch in robot speech will evoke stereotypical evaluations of the robot speaker. To explore this, multiple voices are synthesized from IBM Watson's Text-to-Speech application and then manipulated to have a raised or lowered pitch. Participants are asked to rate these voices on various scales such as competence, trustworthiness, and likeability.
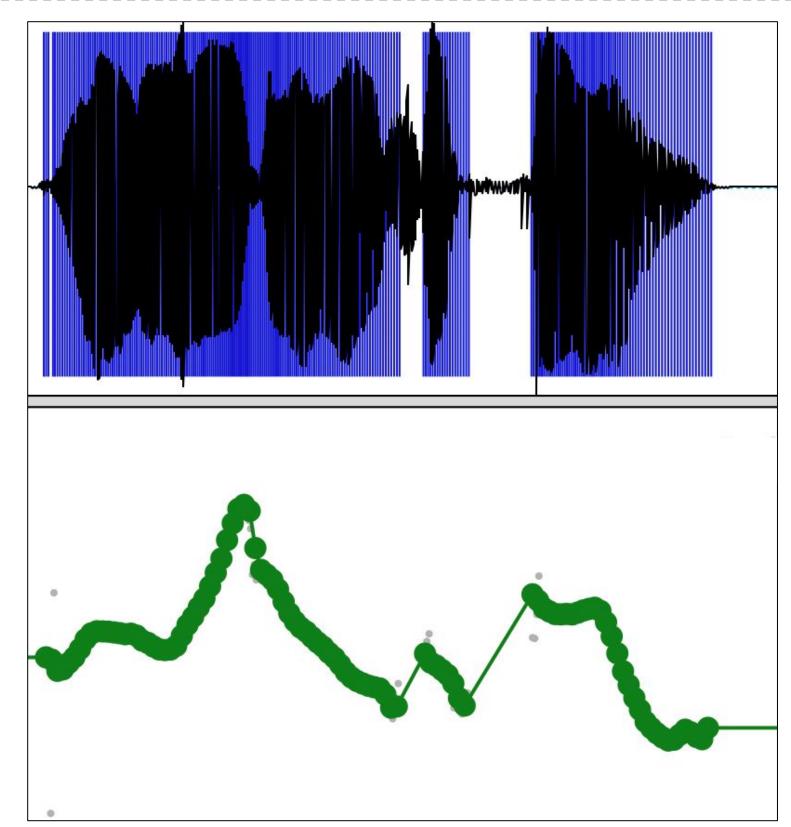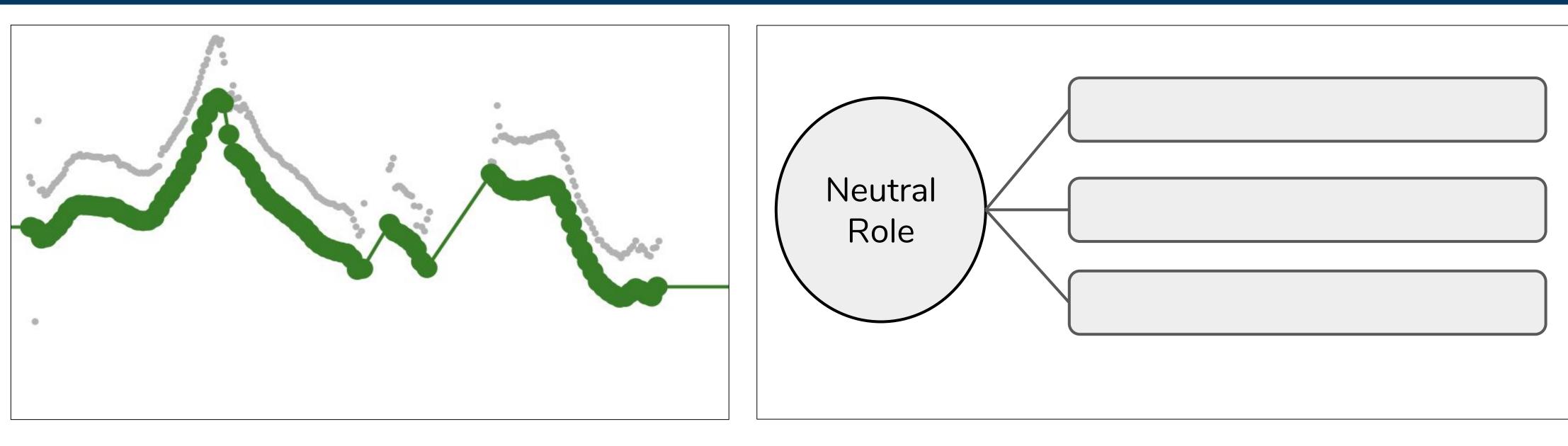

Figure 1: Example Waveform
A view of a speech waveform (top) with its pitch profile (bottom) in Praat.

Approach

Recent research on pitch-based stereotypes in human speech have shown lower-pitched voices are perceived more positively, even across other social stereotypes [1]. Many HRI studies have demonstrated robots are subject to both gender stereotypes [2] and personality stereotypes [3], which are typical of human-human interactions. We present a between-subjects study to investigate whether these strong pitch-based stereotypes are also present in HRI using synthetic speech.

Figure 2: Pitch Profile Modification
The gray represents the original pitch profile from figure 1. The green represents the new pitch profile after subtracting 0.75 ERB in Praat.

Figure 3: Example Configuration
Participants are asked to evaluate three dialogues within each role context. Each dialogue is a unique pitch modification of separate voices.

Methods

Stimuli

The voice stimuli are generated by IBM Watson's Text-to-Speech (TTS) service [4], using the three available American English voices. Figure 1 shows an example speech waveform, as well as its pitch profile, that was generated by IBM Watson's TTS service. The pitch modulation is done in Praat [5] by adjusting (+/-) the pitch by 0.75 equivalent rectangular bandwidths (ERB). As can be seen in Figure 2, this technique allows for the pitch to be effectively modified without affecting the speed of the recording. We obtain nine total voice profiles by raising and lowering the pitch of each voice.

Setup

To investigate pitch-based stereotypes across gender and role stereotypes, we consider:

- Three occupational contexts:
  feminine, neutral, and masculine
- Two voice characteristics:
  feminine and masculine
- Three pitch modifications:
  raised, natural (unaltered), and lowered

Figure 4: Neutral Role Context with Voice Sample
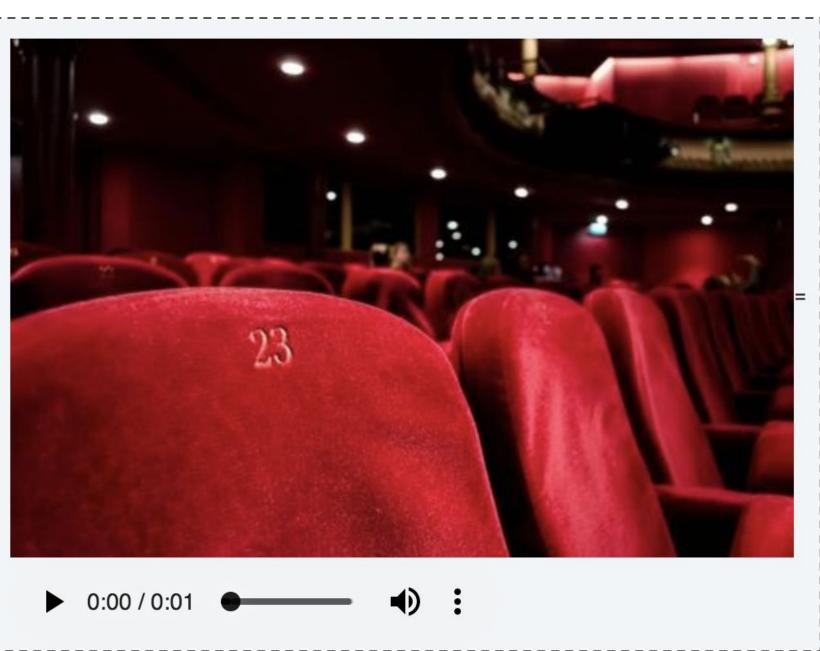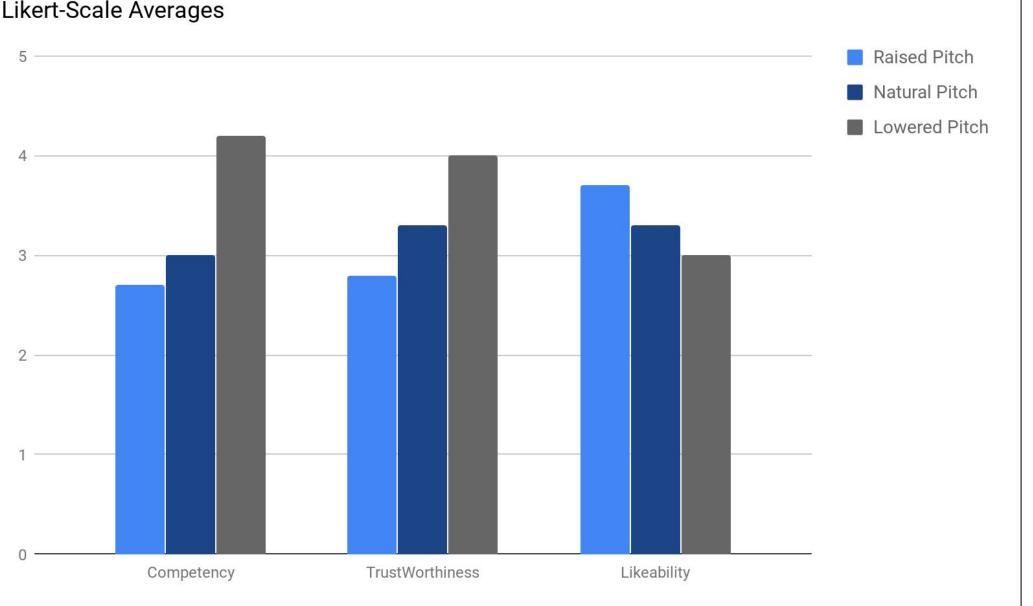We chose an Usher/Theater Attendant as a neutral occupation.

Participants are asked to evaluate three voice profiles, each with a unique pitch modification and dialogue text from separate initial voices, for a particular role context. Figure 3 illustrates an example configuration for the neutral role context. A sample of a role context and voice sample pair that is displayed to the participant can be seen in Figure 4.

Each participant is presented with three total configurations, one for each role context. Participants are given a small questionnaire for each dialogue and asked to evaluate the voice in terms of competency, trustworthiness, and likeability on a Likert scale. Presentation of role contexts, as well as stimuli within, are randomized between participants.

Figure 5: Expected Ratings of Voice Perceptions

Expected Results

Based on previous research which shows preference for lower pitch in human speech [1], we predict lower-pitched voices to score better on competency and trustworthiness evaluations across all role context and voice characteristic combinations. We further predict that while higher pitched voices may not score highly on competency and trustworthiness scales, they will score better in terms of likeability. If our hypotheses are correct, our data would be similar to that of Figure 5.

References

[1] R. C. Anderson and C. A. Klofstad, "Preference for Leaders with Masculine Voices Holds in the Case of Feminine Leadership Roles," PLoS One, vol. 7, (12), 2012.
[2] T. Nomura, "Robots and Gender", Gender and the Genome, vol. 1, no. 1, pp. 18-26, 2017.
[3] B. Tay, Y. Jung and T. Park, "When stereotypes meet robots: The double-edge sword of robot gender and personality in human–robot interaction", Computers in Human Behavior, vol. 38, pp. 75-84, 2014. Available: 10.1016/j.chb.2014.05.014
[4] "Watson Text to Speech", Ibm.com, 2019. [Online]. Available: https://www.ibm.com/watson/services/text-to-speech/. [Accessed: 31- Jul- 2019].
[5] P. Boersma and D. Weenink, "Praat: doing Phonetics by Computer", Fon.hum.uva.nl, 2019. [Online]. Available: http://www.fon.hum.uva.nl/praat/. [Accessed: 31- Jul- 2019].

Acknowledgements