What makes a conversational agent sound trustworthy?

Exploring the role of acoustic-prosodic factors.

Sarah Ita Levitan

Guest Lecture: Advanced Topics in SLP

Columbia University

February 20, 2024

Hello!

- **Current (2020 – present)**

Assistant Prof. of Computer Science, Hunter College

Doctoral faculty of CS and Linguistics, CUNY Graduate Center

- **Previous (2013-2019)**

PhD & Postdoc at Columbia University

Columbia Speech Lab: PI Julia Hirschberg

Outline

- Motivation
- Related work
- Data collection
  - Speech stimuli
  - Crowdsourcing experiment
- Acoustic-prosodic characteristics of trustworthy TTS

Motivation

Motivation

- Trust is essential for effective communication and collaboration
- In human-human interaction AND human-computer interaction
- We understand a great deal about signals of trust in *human* speech
- But have a limited understanding of how humans perceive trustworthiness in *synthesized* speech

What makes a conversational agent sound trustworthy?

Previous work

- Acoustic-Prosodic and Lexical Cues to Deception and Trust: Deciphering How People Detect Lies (Chen & Levitan et al. 2020)
- The sound of trustworthiness: Acoustic-based modulation of perceived voice personality (Belin et al. 2017)

The sound of trustworthiness: Acoustic-based modulation of perceived voice personality (Belin et al. 2017)

- ***How should one say "hello" to be perceived as trustworthy by new listeners?***
- Generate high and low trustworthiness speech stimuli
- Evaluate perception of trustworthiness with online study

Synthesis of Trusthworthy/Untrustworthy stimuli

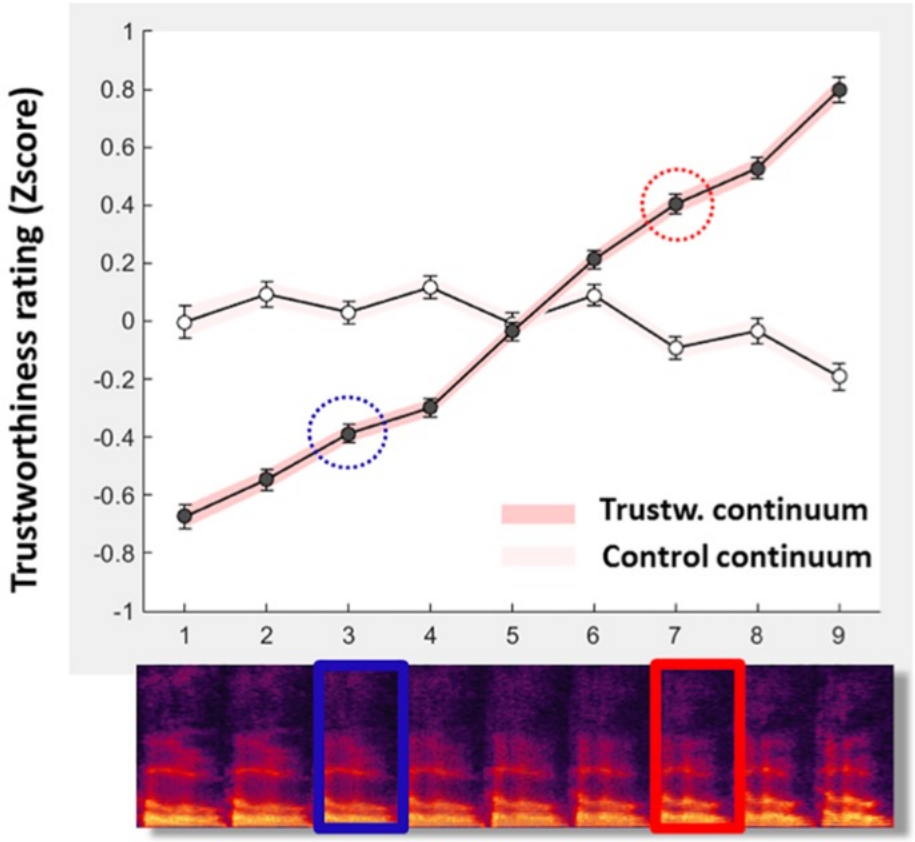
- STRAIGHT toolkit in Matlab
- Decompose natural speech stimulus into 5 parameters:
  - F0, frequency, spectro-temporal density, aperiodicity
- Manipulate and combine parameters
- Synthesize into a novel voice stimulus

Stimuli

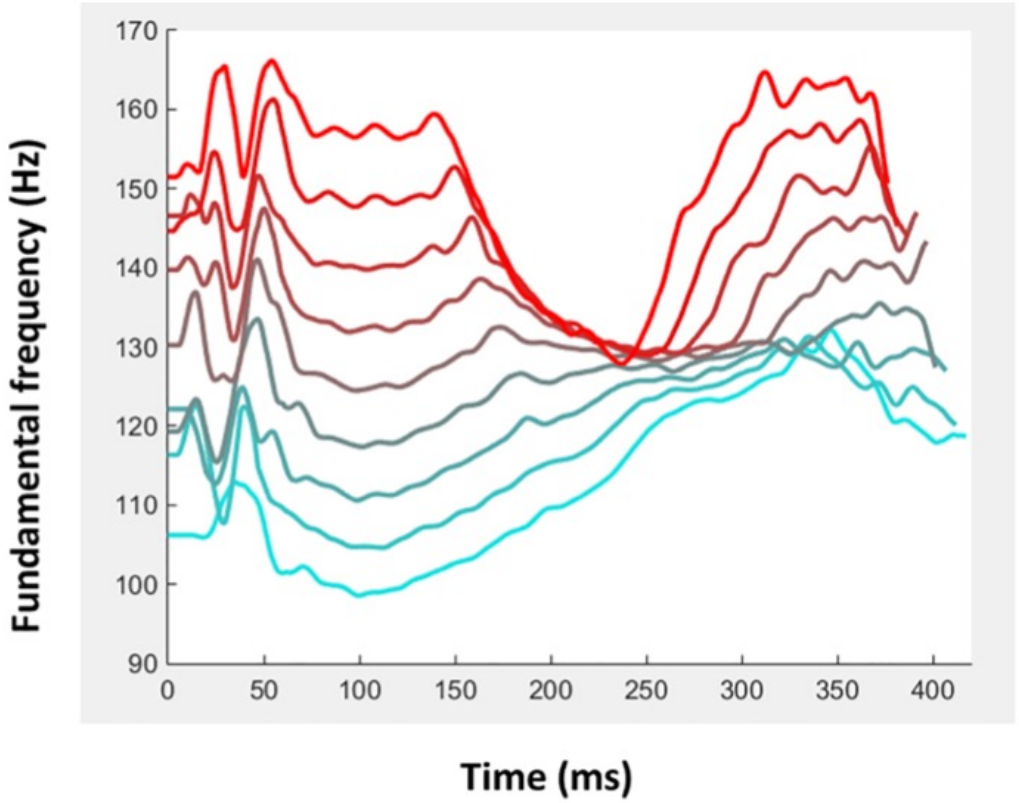
Trust continuum stimuli

Control stimuli

Correlation between acoustics and trust ratings (r=0.99, p=0)



Intonation and perceived trustworthiness

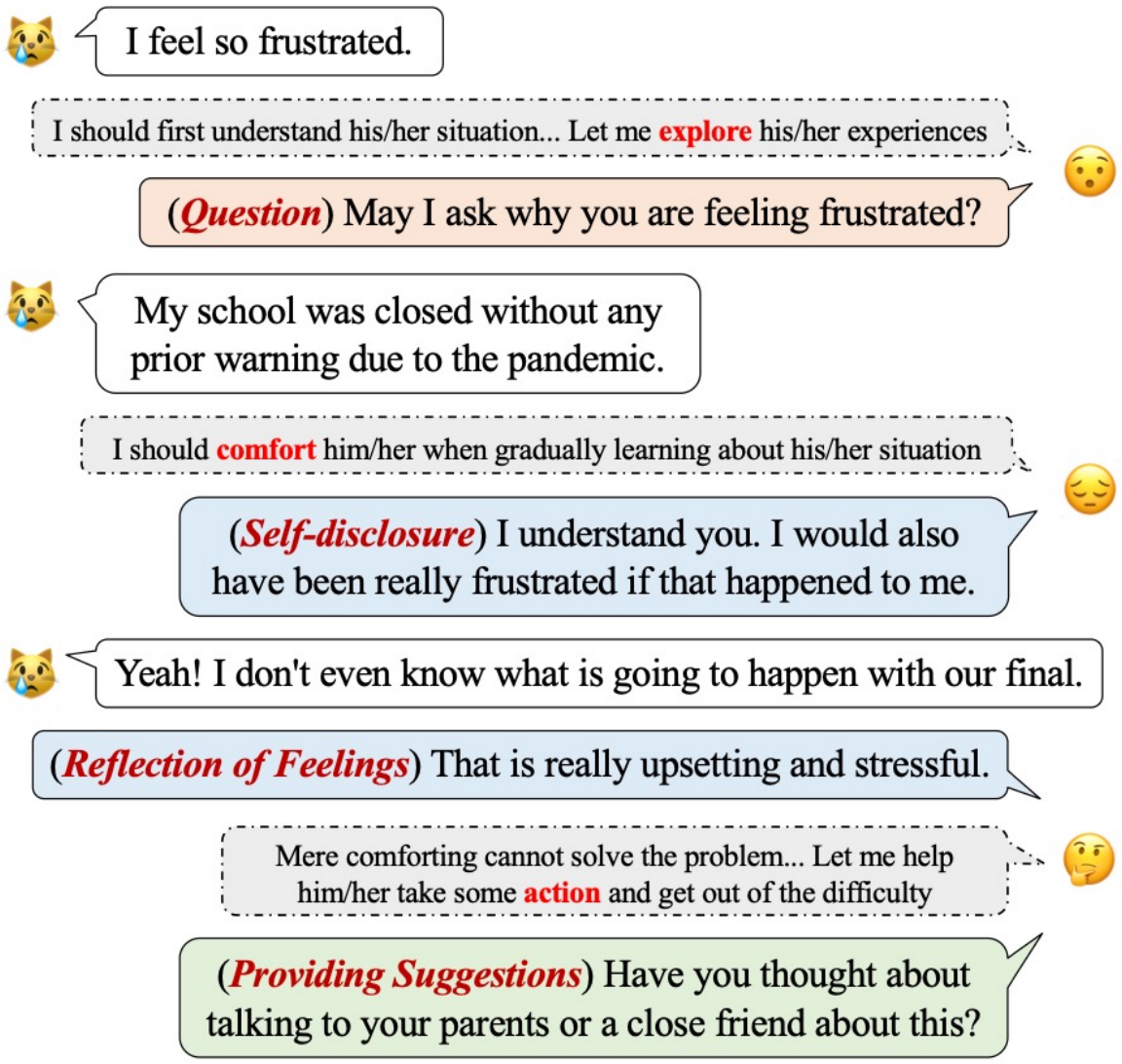


Current Study: Data Collection

- Speech stimuli preparation
- Crowdsourced perception study

Text selection

- Emotional Support Conversations Dataset (Liu et al. 2021)
- 1300 crowdsourced conversations between human help-seeker and virtual supporter
- Application that requires trust and vulnerability from the user
- We select sentences labeled as supporter **questions**



Amazon Polly Neural TTS

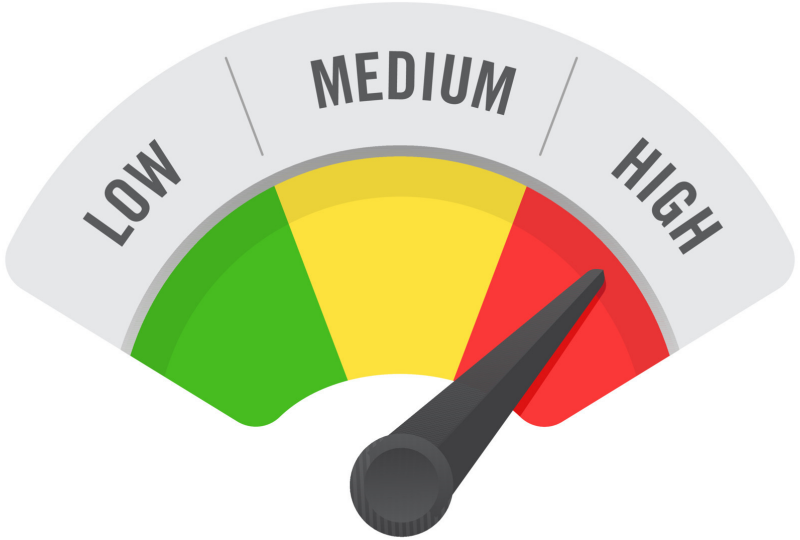
- State-of-the-art, commercial TTS system
- Integrated with dialogue systems and conversational robots
- Supports voice alterations using SSML
- Pre-trained male and female voices

Speech Synthesis Markup Language (SSML)

```
<speak>
<voice name="Joanna"><lang xml:lang="en-US">
<prosody pitch="-27%" rate="95%" volume="+0dB">
Call me Ishmael. <break time="300ms"/> Some years
ago <break time="300ms"/> never mind how long
precisely <break time="300ms"/> having little or
no money in my purse, and nothing particular to
interest me on shore, I thought I would sail
about a little <break time="100ms"/>
and see the watery part of the world.
</prosody></lang></voice>
</speak>
```

Acoustic-prosodic features

- Pitch
- Intensity
- Speaking rate



Total speech stimuli

- 27 prosodic profiles
  - 3 features (pitch, intensity, rate) x 3 settings (low, medium, high)
- 2 voices
  - 1 male (”Matthew”), 1 female (“Joanna”)
- 10 question utterances

- Total: 540 speech samples

Mean feature values

Level	Intensity		Pitch		Speaking rate	
	F	M	F	M	F	M
High	57.2	55	203	131	294	321
Med	52	51	163	110	236	256
Low	49	47	131	103	192	205

Examples

- Low pitch, intensity, speaking rate

- Medium pitch, intensity, speaking rate

- High pitch, intensity, speaking rate

Crowdsourced Perception Study

- Listen to 20 audio clips
- Rate speaker traits with 5-point Likert scale
  - Trustworthy, lively, empathetic, respectful, cold, engaging
- Quality control: transcription task
- Listener traits:
  - Ten Item Personality Inventory (TIPI)
  - Gender

Crowdsourced Perception Study

- 135 participants (71 F, 63 M)
- Each audio sample was rated by 5 unique raters
- 2700 judgments of 540 speech stimuli
- All judgments are z-normalized by rater

Inter-Annotator Agreement

- Krippendorff's alpha

	trustworthy	**lively**	**natural**	**boring**	**empathetic**	**respectful**	**cold**	**engaging**
all raters	0.21	0.18	0.17	0.2	0.2	0.18	0.22	0.2
F raters	0.13	0.08	0.08	0.07	0.1	0.1	0.12	0.1
M raters	0.17	0.13	0.14	0.17	0.14	0.15	0.16	0.13

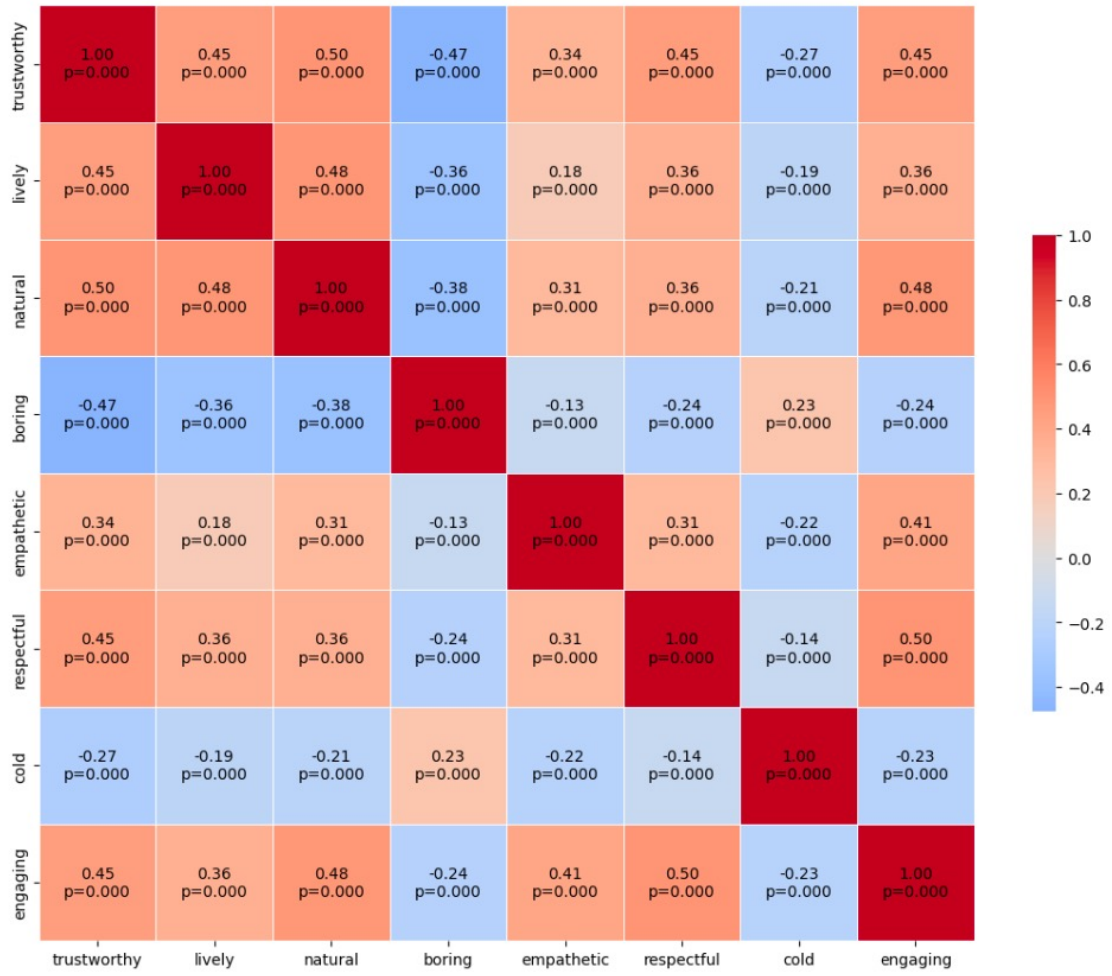
Average ratings per trait

	trustworthy	**lively**	**natural**	**boring**	**empathetic**	**respectful**	**cold**	**engaging**
all raters	3.66	3.66	3.48	2.6	3.34	3.7	3	3.6
F raters	3.69	3.7	3.6	2.5	3.3	3.6	3	3.6
M raters	3.66	3.6	3.4	2.6	3.3	3.7	3	3.5

Questions

- How do raters define trustworthiness in terms of other speaker traits?
  - Lively, empathetic, respectful, cold, engaging
- What are the acoustic-prosodic characteristics of trustworthy speech?
  - And other speaker traits?
- How do listener characteristics (gender, personality) affect their perception of trustworthiness and other speaker traits?

Correlation analysis of speaker attributes



Acoustic-prosodic characteristics of trustworthy TTS

Intensity	Pitch	Rate	Gender	Avg Rating
medium	medium	medium	M	4.1
high	high	medium	F	4
low	high	low	M	3.95
medium	high	medium	M	3.91
medium	high	low	F	3.9
low	low	high	M	3.2
high	low	high	M	3.15
medium	medium	high	M	3.15
high	low	high	F	3
high	medium	high	M	2.9

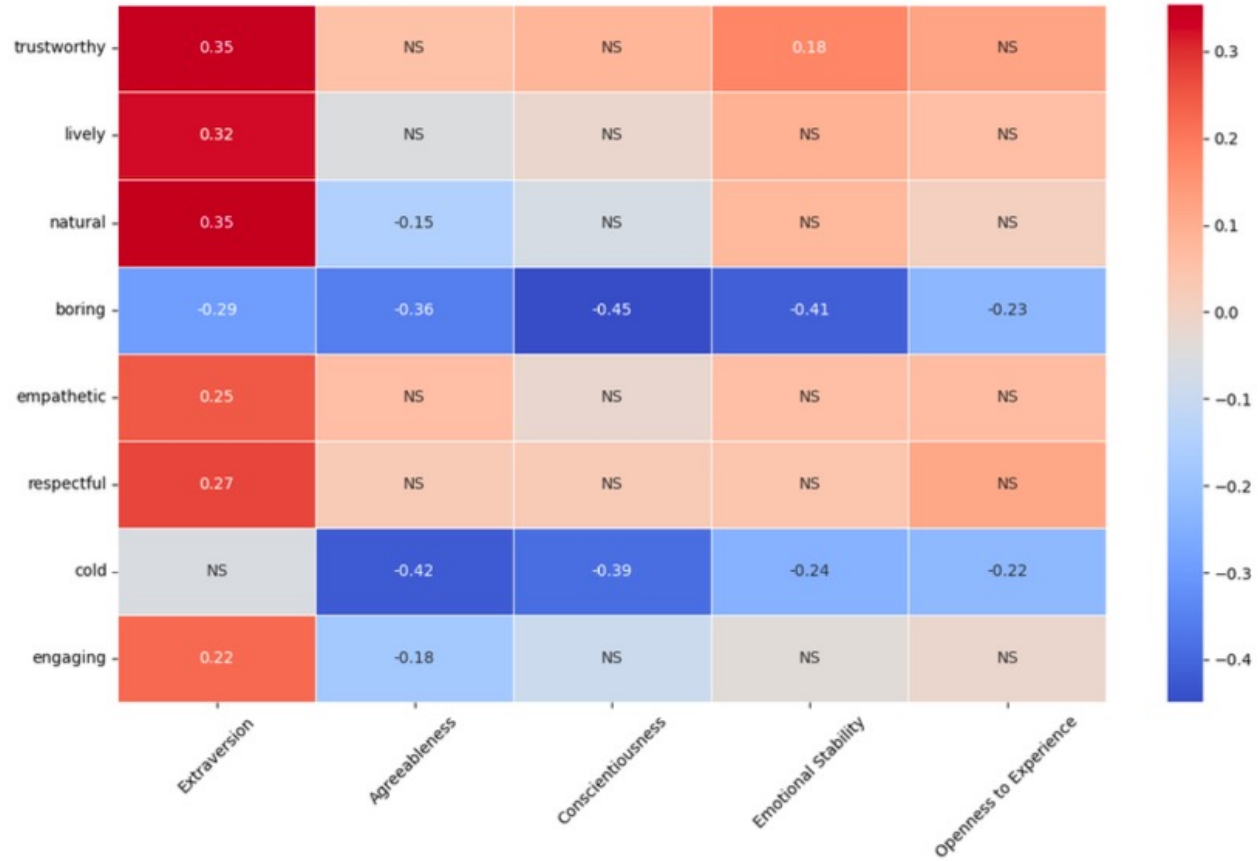
GLS Regression Analysis

Features	trustworthy	
	r	p
intensity low	-0.13	0
intensity medium	0.31	0
intensity high	-0.17	0
pitch low	-0.17	0
pitch medium		
pitch high	0.29	0
speaking rate low	0.3	0
speaking rate medium	0.4	0
speaking rate high		

How does listener gender affect their perception?

- Female listeners were more likely to perceive speakers as natural (r=-0.18) and empathetic (*r*=-0.07)
- Male listeners were more likely to perceive speakers as boring (*r*=0.11)

How does listener personality affect their perception?



Summary

- Crowdsourced perception study of trustworthy synthesized speech
- Identified specific patterns of synthesized speech associated with perceived trustworthiness
- Listener gender and personality traits may affect perception

- Next steps: explore the role of lexical factors
  - Dialogue act
  - Politeness
  - Complexity

Thank you!

- Yuwen Yu: PhD student, CUNY Graduate Center
- Ghazanfar Shahbaz: Previous undergraduate student, Hunter College
- Funding: NSF EAGER

Questions?