“It’s not a representation of me”: Examining Accent Bias and Digital Exclusion in Synthetic AI Voice Services

Shira Michel*, Sufi Kaur*, Sarah Elizabeth Gillespie*, Jeffrey Gleason*, Christo Wilson*, Avijit Ghosh+

* Northeastern University
+ Hugging Face 🤗 and University of Connecticut

ACM Conference on Fairness, Accountability, and Transparency (FAccT) 2025

Background

There is a broader issue of accent bias and digital exclusion faced by those with diverse accents.

Image credits: thetimes.com, peopleofcolorintech.com, motherjones.com

Background

Why Do Cartoon Villains Speak in Foreign Accents?

Children's shows often use non-standard dialects to voice the "bad guys," sending a dangerous message to kids about diversity.

By Isabel Fattal

wavebreakmedia / Skylines / Porfang / Shutterstock / Disney / Zak Bickel / The Atlantic

Poll brands Scouse accent one of the least attractive - tell us why it's them who are talking rubbish

A study by YouGov saw the Scouse accent ranked as one of the least attractive accents in the British Isles

NEWS By Emilia Bona Content Editor

13:42, 17 Aug 2015 | Updated 11:47, 16 Nov 2015

YouGov | yougov.com November 27-28, 2014

Which accents are the most and least attractive?

Media culture shapes identity and societal values that influence how we perceive morality and power.

Image credits: theatlantic.com, www.mirror.co.uk

Background

Accent ≠ dialect [1].

Image credits: thevarsity.ca

[1] Markl and Lai. 2023. ISCA.

Background

Accent ≠ dialect [1].

Accents are a **key part of identity**, *everyone* has one [1].

Image credits: thevarsity.ca

[1] Markl and Lai. 2023. ISCA.

Background

Accent ≠ dialect [1].

Accents are a **key part of identity**, *everyone* has one [1].

Familiar accents **improve listener comprehension and cognitive processing** [2,3,4].

Image credits: thevarsity.ca

[1] Markl and Lai. 2023. ISCA. [2] Adank et al. 2009. Journal of Experimental Psychology: Human perception and performance 35. [3] Njie et al. 2023. Memory & Cognition 51. [4] Perry et al. 2018. Psychonomic bulletin & review 25.

Background

Accent ≠ dialect [1].

Accents are a **key part of identity**, *everyone* has one [1].

Familiar accents **improve listener comprehension and cognitive processing** [2,3,4].

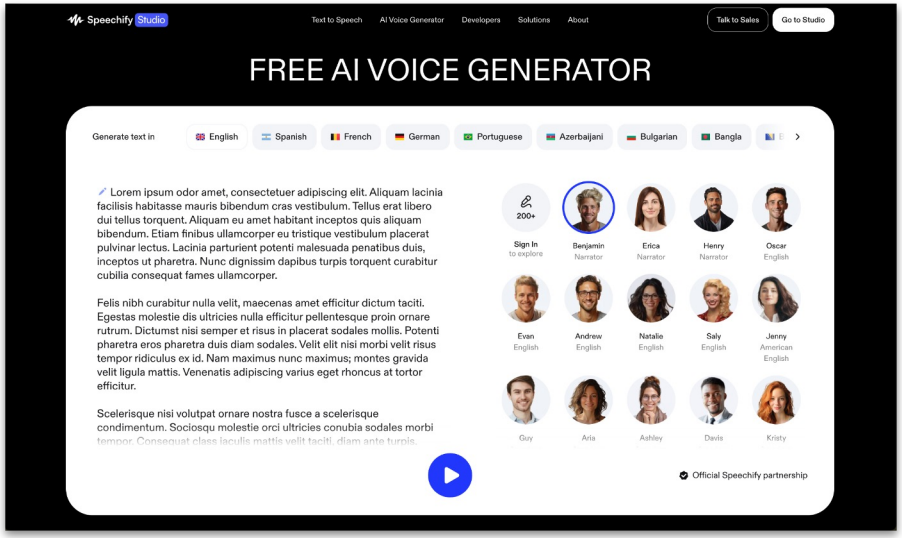
Our work focuses on **issues that may stem from low-accuracy systems** for the English language.

Image credits: thevarsity.ca

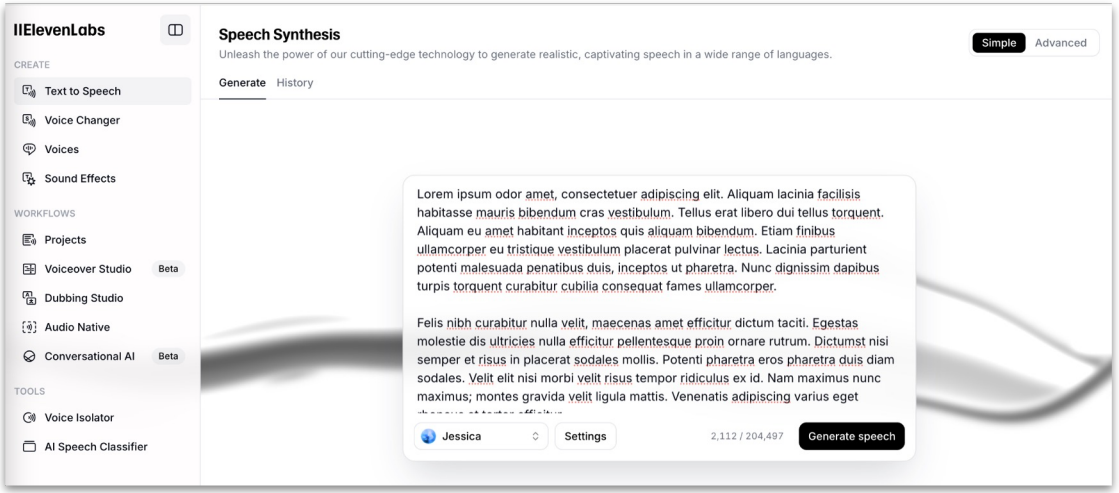
[1] Markl and Lai. 2023. ISCA. [2] Adank et al. 2009. Journal of Experimental Psychology: Human perception and performance 35. [3] Njie et al. 2023. Memory & Cognition 51. [4] Perry et al. 2018. Psychonomic bulletin & review 25.

Study Overview

Speechify



IIElevenLabs

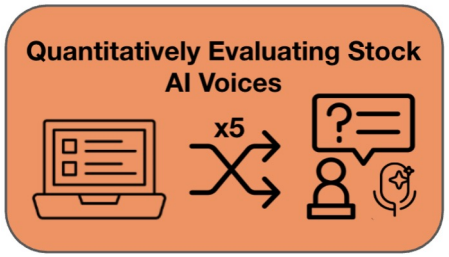


*We investigated **American, Australian, African, British, and Indian English-language accents** as labeled by these services*.*

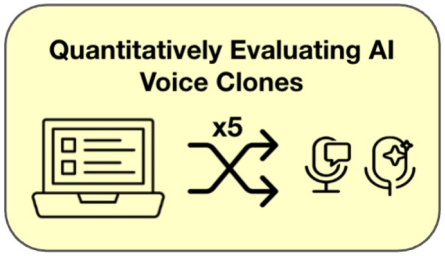
**As of May 2024, ElevenLabs offered only these five accents while Speechify provided these same accents as options.*
Image credits: speechify.com, elevenlabs.io

Study Overview

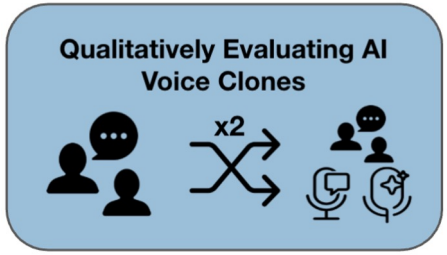
Study 1



Study 2

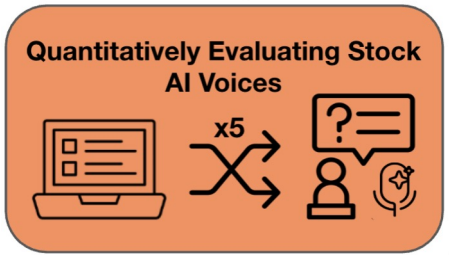


Study 3

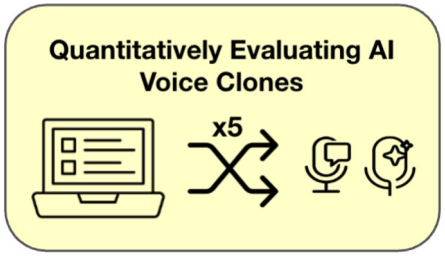


Study Overview

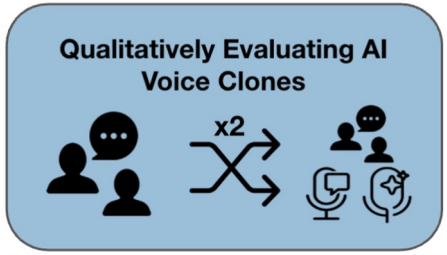
Study 1



Study 2



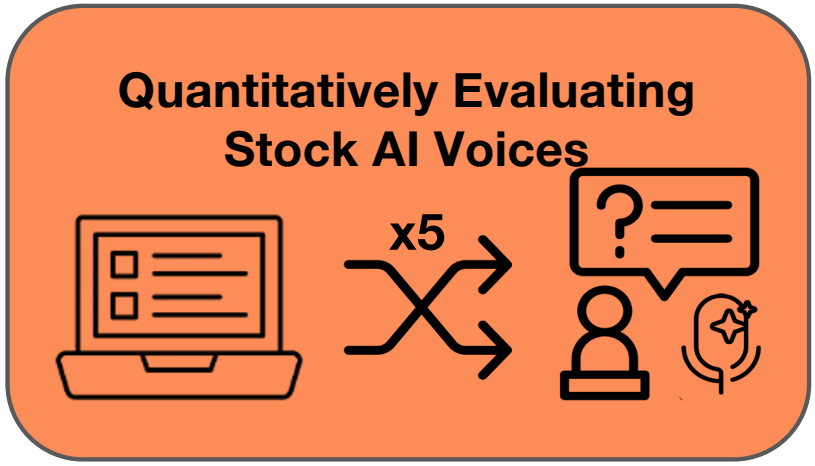
Study 3



How well do synthetic AI voice services capture and represent diverse accents?

Study 1

Study 1

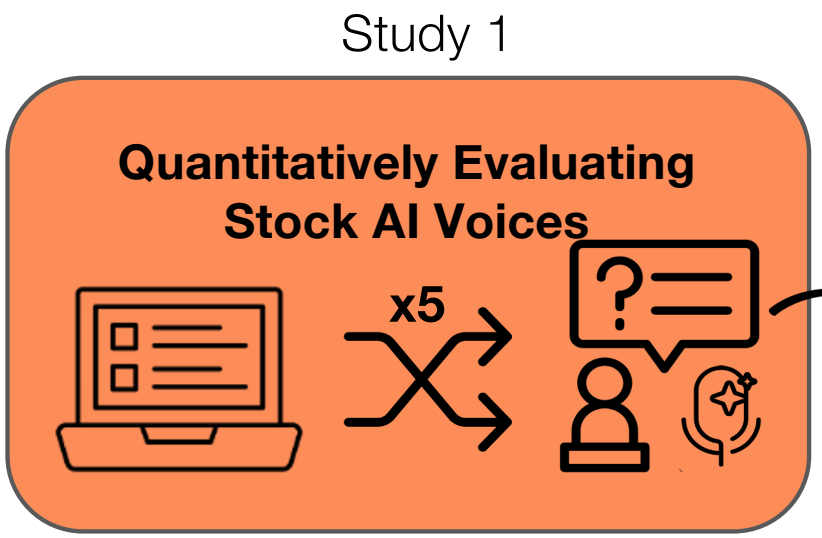


N = 250 (diverse accents and pitches*)

Each participant listens to **5 stock AI voices**

** We define pitches as having high- or low-pitched voices.*
[5] Viswanathan and Viswanathan. 2005. Computer speech & language 19.

Study 1



N = 250 (diverse accents and pitches*)

Each participant listens to **5 stock AI voices**

Labeling task:

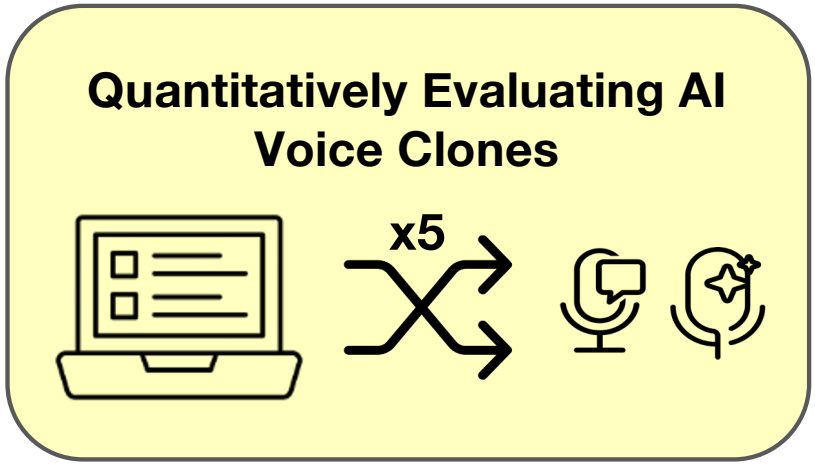
Voice Quality (Modified Mean Opinion Score [5])**:** *How would you describe the naturalness of the audio?*

Voice Persona: *What accent do you think the speaker of the voice identifies with?*

** We define pitches as having high- or low-pitched voices.*
[5] Viswanathan and Viswanathan. 2005. Computer speech & language 19.

Study 2

Study 2



N = 250 (diverse accents and pitches*)

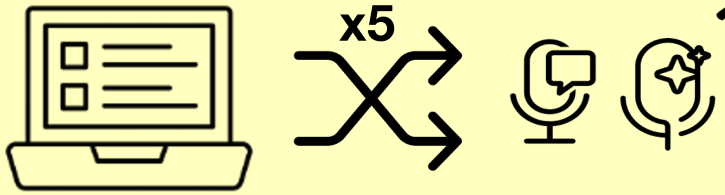
Each participant listens to **5 pairs of recorded and clones voices**

**Pitches = high- or low-pitched voices*

Study 2

Study 2

Quantitatively Evaluating AI Voice Clones



N = 250 (diverse accents and pitches*)

Each participant listens to **5 pairs of recorded and clones voices**

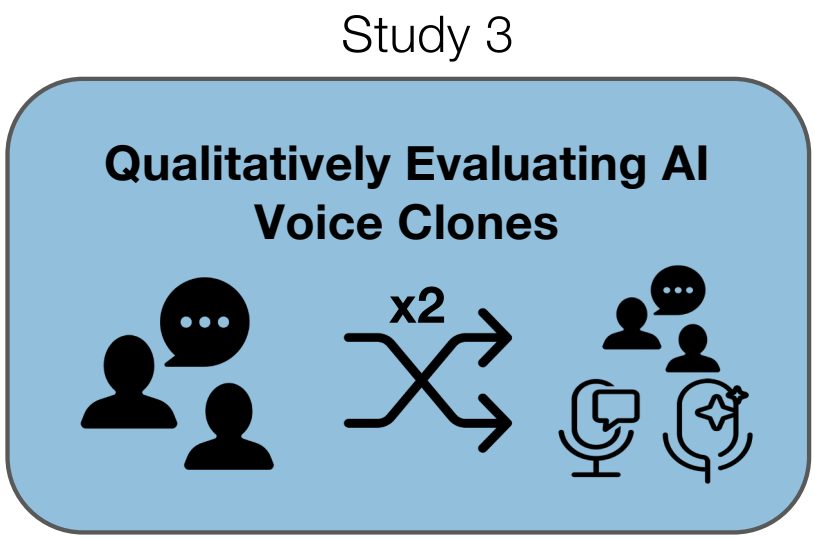
Labeling task:

Naturalness of Voice:** *How natural (i.e., human-sounding) is AUDIO 1?*

Accent Mimicry Accuracy (AMA):** *How well do you think AUDIO 2 emulated the speech style of AUDIO 1?*

**Pitches = high- or low-pitched voices*
*** Please see our paper for details on these metrics*

Study 3

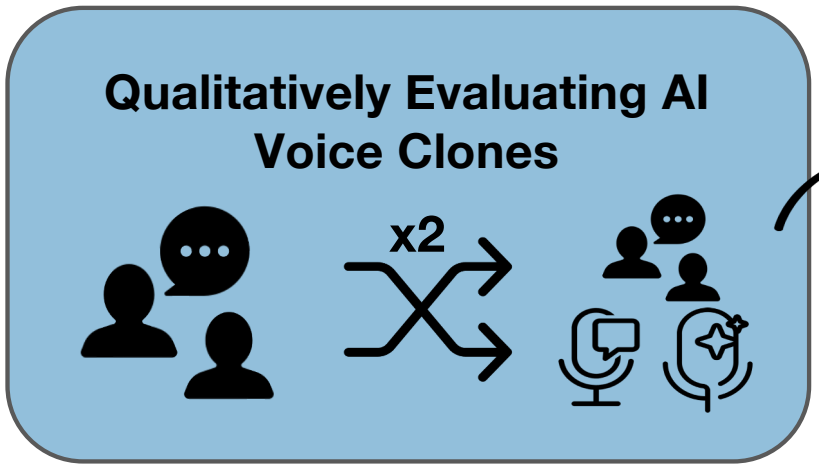


N = 26 (diverse accents and pitches*)

**Pitches = high- or low-pitched voices*

Study 3

Study 3



N = 26 (diverse accents and pitches*)

Interview Guide:

a) Before Listening Questions

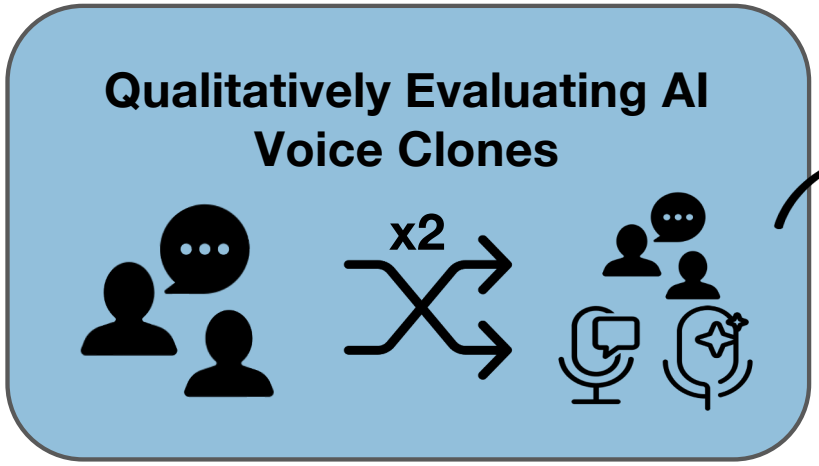
Relationship with Voice and Accent

Familiarity with AI-generated Speech

**Pitches = high- or low-pitched voices*

Study 3

Study 3



N = 26 (diverse accents and pitches*)

Each participant listens to **2 pairs of their recorded and clones voices**

Interview Guide:

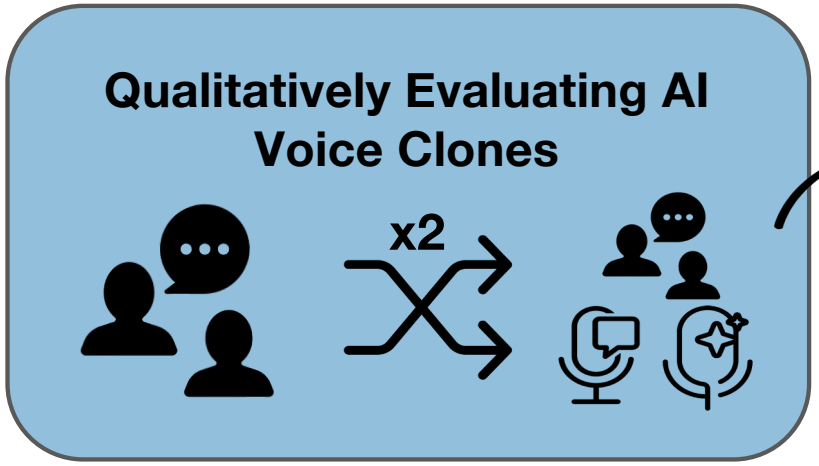
a) Before Listening Questions
Relationship with Voice and Accent

Familiarity with AI-generated Speech

**Pitches = high- or low-pitched voices*

Study 3

Study 3



N = 26 (diverse accents and pitches*)

Each participant listens to **2 pairs of their recorded and clones voices**

Interview Guide:

a) Before Listening Questions
Relationship with Voice and Accent

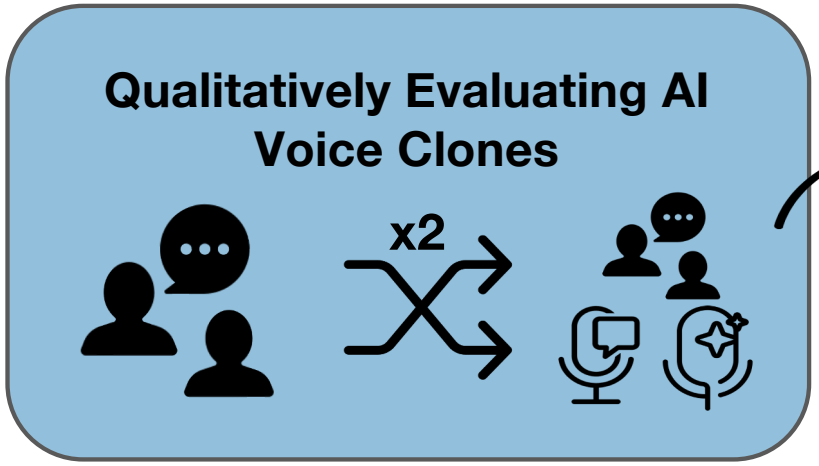
Familiarity with AI-generated Speech

b) Evaluation
Recreation and Accent Mimicry Quality (AMQ):** *How well do you think the generated voice created your voice?*

**Pitches = high- or low-pitched voices*
*** Please see our paper for details on these metrics*

Study 3

Study 3



N = 26 (diverse accents and pitches*)

Each participant listens to **2 pairs of their recorded and clones voices**

Interview Guide:

a) Before Listening Questions
Relationship with Voice and Accent

Familiarity with AI-generated Speech

b) Evaluation
Recreation and Accent Mimicry Quality (AMQ):** *How well do you think the generated voice created your voice?*

c) After Listening Questions
Reactions

Concerns

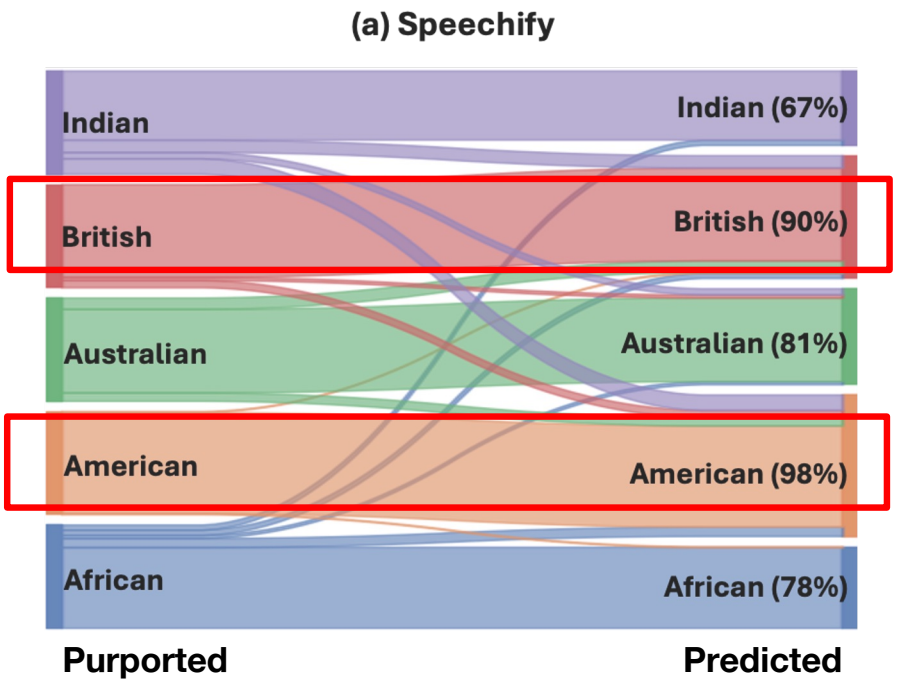
**Pitches = high- or low-pitched voices*
*** Please see our paper for details on these metrics*

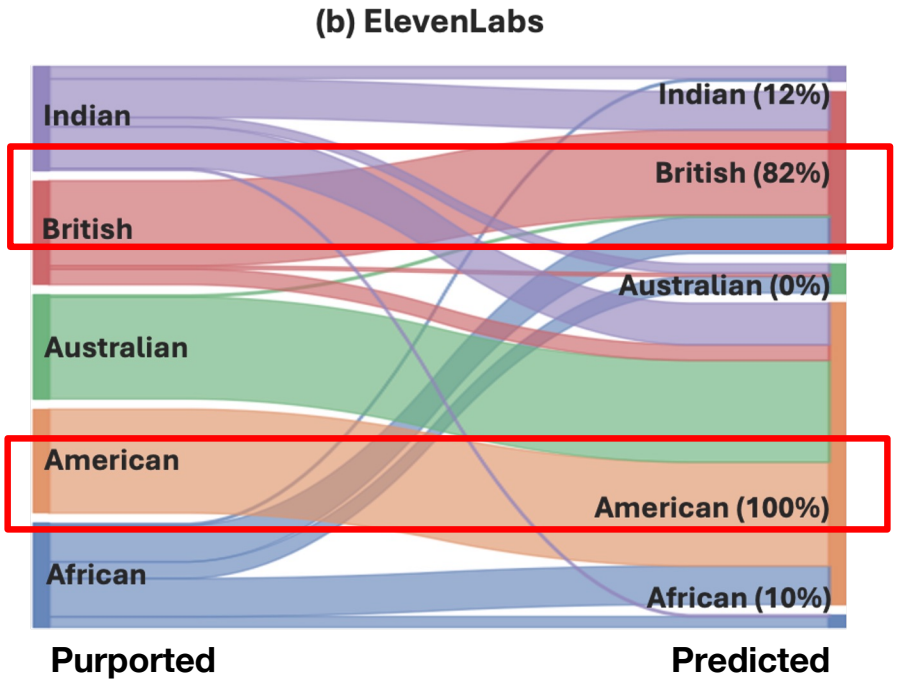
Key Findings

(1) *There are* ***disagreements in accent classification****, but* ***voices generated by ElevenLabs received higher ratings*** *for quality and cloning accuracy.*

Key Findings

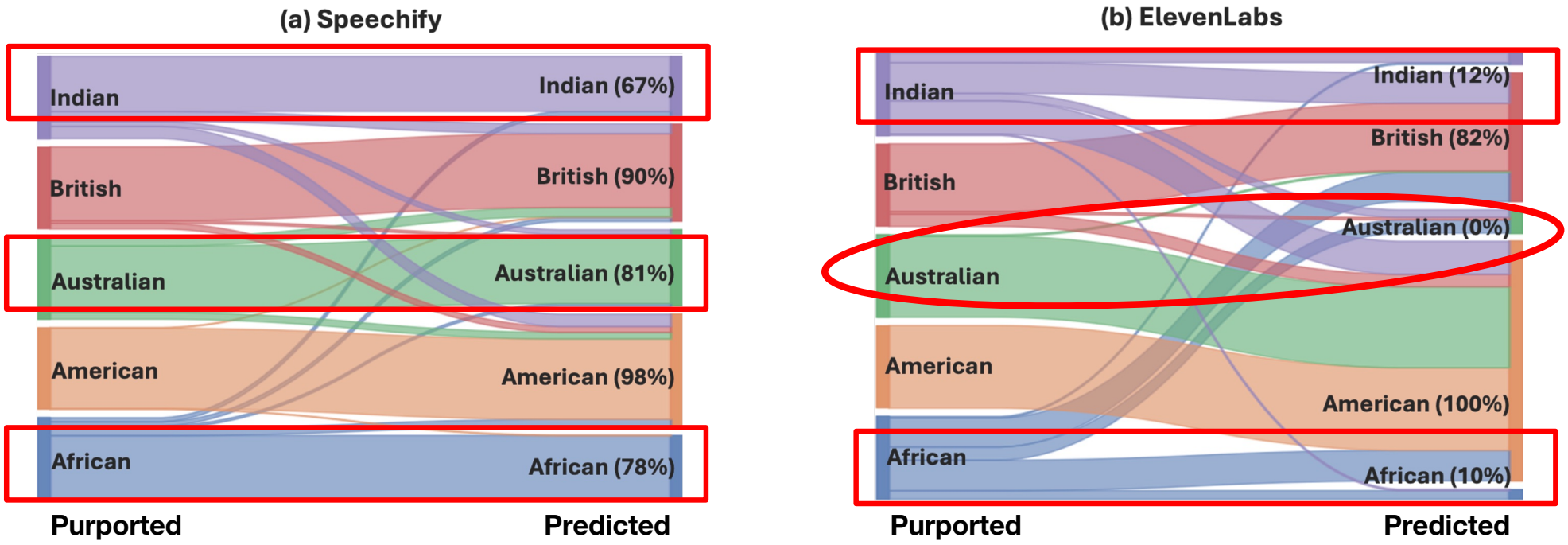
***(1)** There are **disagreements in accent classification**, but **voices generated by ElevenLabs received higher ratings** for quality and cloning accuracy.*





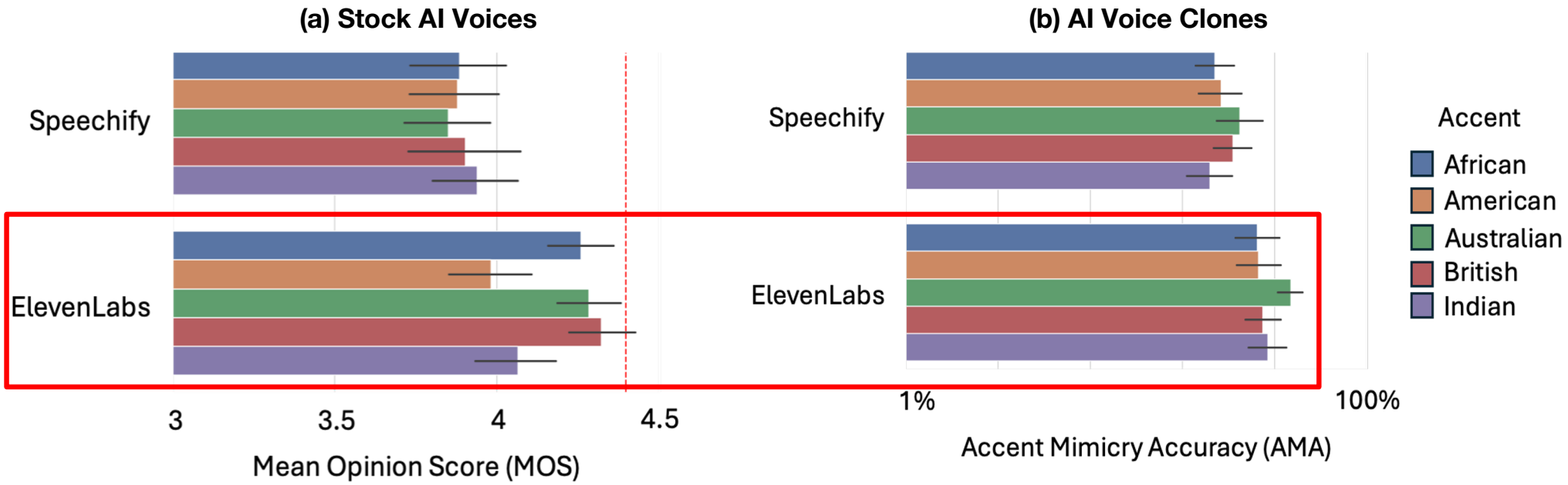
Key Findings

***(1)** There are **disagreements in accent classification**, but **voices generated by ElevenLabs received higher ratings** for quality and cloning accuracy.*



Key Findings

***(1)** There are **disagreements in accent classification**, but **voices generated by ElevenLabs received higher ratings** for quality and cloning accuracy.*



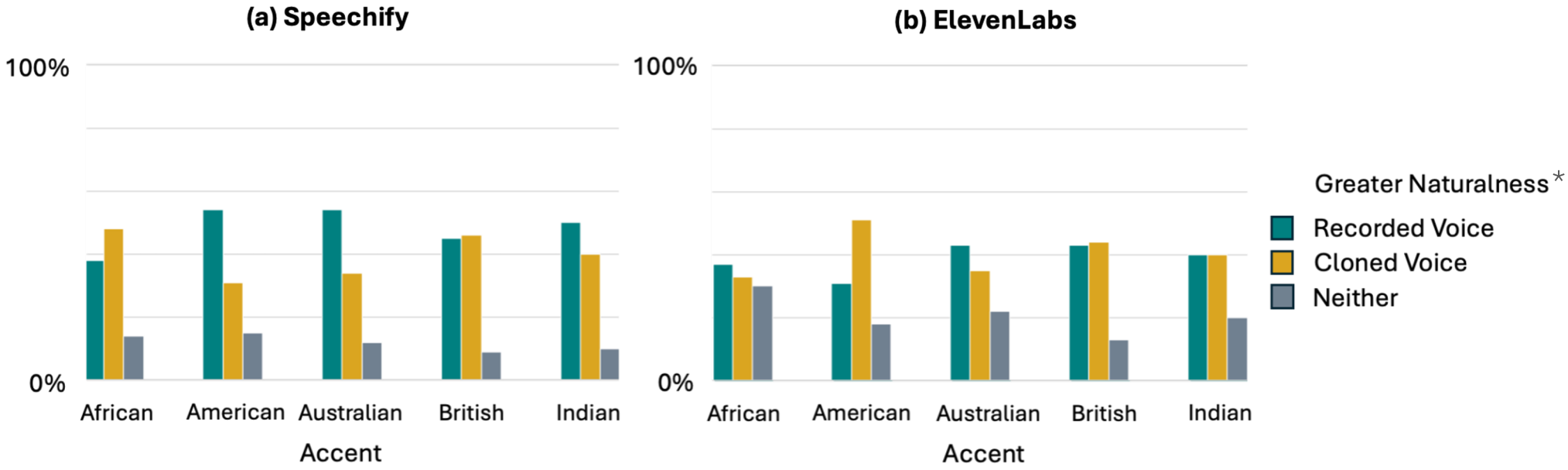
Excellent quality reached at baseline = 4.4 (Twilio and WayWithWords)

Key Findings

*(2) Users **struggle to discern** between human and synthetic voices.*

Key Findings

*(2) Users **struggle to discern** between human and synthetic voices.*



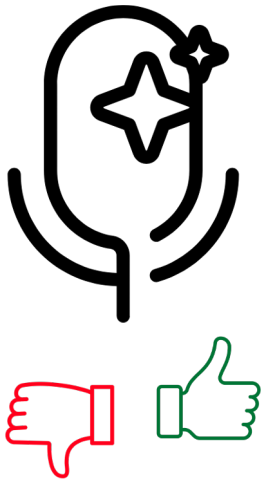
*Naturalness (i.e., human-sounding)

Key Findings

(3) *Accent mimicry accuracy alone* ***does not determine user satisfaction.***

Key Findings

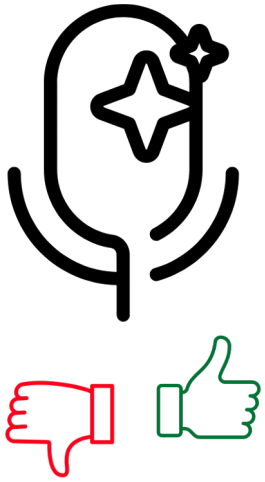
(3) *Accent mimicry accuracy alone* ***does not determine user satisfaction.***



66.7% of responses **preferred the voice clone from ElevenLabs**, but 52% **favored a robotic AI voice clone.**

Key Findings

(3) *Accent mimicry accuracy alone* ***does not determine user satisfaction.***



66.7% of responses **preferred the voice clone from ElevenLabs**, but 52% **favored a robotic AI voice clone.**

"*...I can definitely think of use cases, but like none that I'd like participate in myself like,* ***unless I became like a con artist****...*" - P25

Key Findings

(4) *There are* ***deeper emotional and cultural dimensions*** *of how users interact with AI-generated versions of their own voices.*

Key Findings

(4) *There are* ***deeper emotional and cultural dimensions*** *of how users interact with AI-generated versions of their own voices.*

53.8% of responses noted **American and British accents overwhelming define the voice standard for AI systems** and excluded other accents.

Key Findings

(4) *There are* ***deeper emotional and cultural dimensions*** *of how users interact with AI-generated versions of their own voices.*

53.8% of responses noted **American and British accents overwhelming define the voice standard for AI systems** and excluded other accents.

"...*just having those [American and British accents]. We just know that. Well,* ***maybe they were not meant for us. They were building them with some other people in mind***." - P18

Key Findings

(4) *There are* ***deeper emotional and cultural dimensions*** *of how users interact with AI-generated versions of their own voices.*

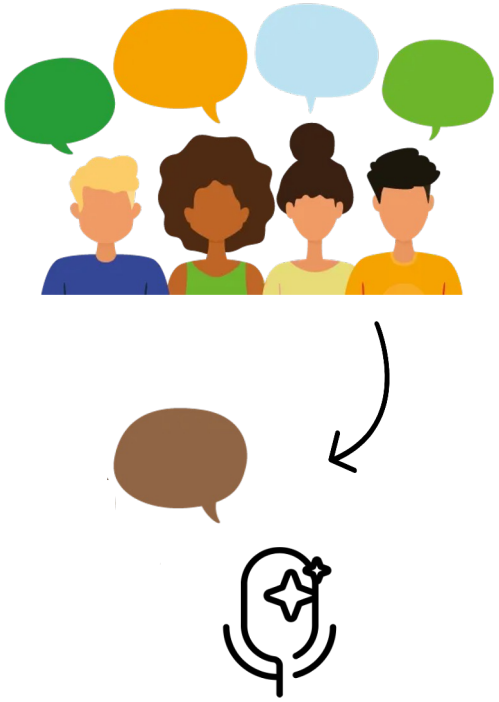


Image credits: crossroadsspeechtherapy.com

"*...****it sounded extremely typical of how, like, an Indian person speaks****. I definitely don't think that's how I speak.* ***It's not a representation of me****..." - P10*

Key Findings

***(4)** There are **deeper emotional and cultural dimensions** of how users interact with AI-generated versions of their own voices.*

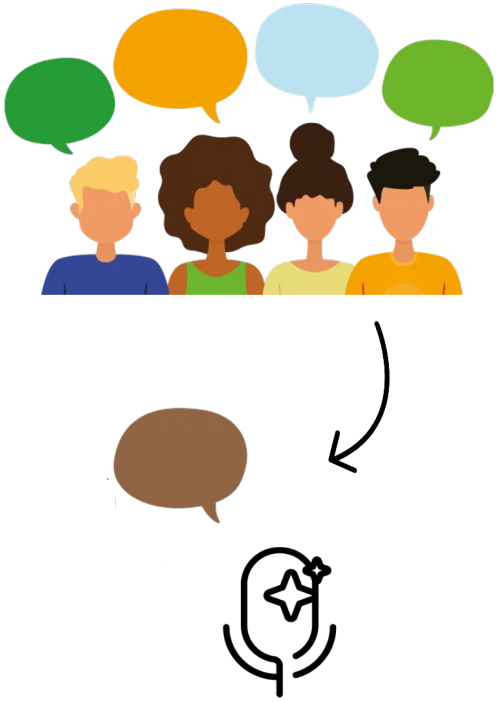


Image credits: crossroadsspeechtherapy.com

"*...**it sounded extremely typical of how, like, an Indian person speaks**. I definitely don't think that's how I speak. **It's not a representation of me**...*" - P10

"***...**it sounded **more formal, more refined**... it's an accent I would **aspire** to speak with...*" - P15

Key Takeaways

***(1)** There are **disagreements in accent classification**, but **voices generated by ElevenLabs received higher ratings** for quality and cloning accuracy.*

↳ The need for **more accurate and culturally informed** accent designations.

Key Takeaways

(1) There are disagreements in accent classification, but voices generated by ElevenLabs received higher ratings for quality and cloning accuracy.

↳ The need for more accurate and culturally informed accent designations.

(2) *Users* ***struggle to discern*** *between human and synthetic voices.*

↳ The importance of **human or mechanically distinguishable** AI-generated speech.

Key Takeaways

(1) There are disagreements in accent classification, but voices generated by ElevenLabs received higher ratings for quality and cloning accuracy.

↳ The need for more accurate and culturally informed accent designations.

(2) Users struggle to discern between human and synthetic voices.

↳ The importance of human or mechanically distinguishable AI-generated speech.

(3)** Accent mimicry accuracy alone **does not determine user satisfaction.

↳ The **intricate connection** between technical performance and user acceptance.

Key Takeaways

(1) There are disagreements in accent classification, but voices generated by ElevenLabs received higher ratings for quality and cloning accuracy.

↳ The need for more accurate and culturally informed accent designations.

(2) Users struggle to discern between human and synthetic voices.

↳ The importance of human or mechanically distinguishable AI-generated speech.

(3) Accent mimicry accuracy alone does not determine user satisfaction.

↳ The intricate connection between technical performance and user acceptance.

***(4)** There are **deeper emotional and cultural dimensions** of how users interact with AI-generated versions of their own voices.*

↳ Reactions **tied to identity** reveal **tensions** between authenticity and aspiration.

Key Takeaways

(1) *There are* ***disagreements in accent classification****, but* ***voices generated by ElevenLabs received higher ratings*** *for quality and cloning accuracy.*

(2) *Users* ***struggle to discern*** *between human and synthetic voices.*

(3) *Accent mimicry accuracy alone* ***does not determine user satisfaction.***

(4) *There are* ***deeper emotional and cultural dimensions*** *of how users interact with AI-generated versions of their own voices.*

Key Takeaways

(1) *There are* ***disagreements in accent classification****, but* ***voices generated by ElevenLabs received higher ratings*** *for quality and cloning accuracy.*

(2) *Users* ***struggle to discern*** *between human and synthetic voices.*

(3) *Accent mimicry accuracy alone* ***does not determine user satisfaction.***

(4) *There are* ***deeper emotional and cultural dimensions*** *of how users interact with AI-generated versions of their own voices.*

Developers: Improve Technical Performance

Key Takeaways

(1) *There are* ***disagreements in accent classification****, but* ***voices generated by ElevenLabs received higher ratings*** *for quality and cloning accuracy.*

(2) *Users* ***struggle to discern*** *between human and synthetic voices.*

(3) *Accent mimicry accuracy alone* ***does not determine user satisfaction.***

(4) *There are* ***deeper emotional and cultural dimensions*** *of how users interact with AI-generated versions of their own voices.*

Developers: Improve Technical Performance

Policymakers: Address Misuse

Key Takeaways

(1) *There are* ***disagreements in accent classification****, but* ***voices generated by ElevenLabs received higher ratings*** *for quality and cloning accuracy.*

(2) *Users* ***struggle to discern*** *between human and synthetic voices.*

(3) *Accent mimicry accuracy alone* ***does not determine user satisfaction.***

(4) *There are* ***deeper emotional and cultural dimensions*** *of how users interact with AI-generated versions of their own voices.*

Developers: Improve Technical Performance

Policymakers: Address Misuse

Employers: Ensure Fair Treatment of Accent Diversity

Northeastern University UCONN UNIVERSITY OF CONNECTICUT

For more information, please refer to our paper.

“It’s not a representation of me”: Examining Accent Bias and Digital Exclusion in Synthetic AI Voice Services


Shira Michel


Sufi Kaur


Sarah Elizabeth Gillespie


Jeffrey Gleason


Christo Wilson


Avijit Ghosh